

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:28 ; Search time 85.6067 Seconds

(without alignments)
3062.893 Million cell updates/sec

Title: US-10-657-280-1
Perfect score: 4907
Sequence: 1 MFLKLRQRYTCLSHRYGL.....SLKQLAVFTDSKNTGRQLK 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4907	100.0	928	4 ABR61376	Abt61376 Human Glc
2	4907	100.0	928	7 ADD27812	Add27812 Human Glc
3	4907	100.0	928	7 ABR01488	Abv01488 Human Glc
4	4907	100.0	928	7 ABR01537	Abv01537 Human Glc
5	4886	99.8	1256	5 AAE25290	Aae25290 Human nuc
6	4630	94.4	1199	7 ADD27810	Add27810 Soluble h
7	4630	94.4	1199	7 ABR01487	Abv01487 N-acetylg
8	4630	94.4	1199	7 ABR01536	Abv01536 N-acetylg
9	4500	91.7	1196	5 AAE25294	Aae25294 Human nuc
10	4468	91.1	846	6 ADA54824	Ada54824 Human pro
11	4391	89.5	1459	6 ABR07381	Abv07381 Human pro
12	3894	79.4	908	4 ABR61384	Abt61384 Murine Gl
13	3894	79.4	908	7 ADD27817	Add27817 GlcNAc-ph
14	3894	79.4	908	7 ABR01491	Abv01491 Mouse pro
15	3894	79.4	908	7 ABR01540	Abv01540 Mouse pro
16	1900	38.7	367	4 AAM16457	Aam16457 Peptide #
17	1900	38.7	367	4 AAM28952	Aam28952 Peptide #
18	1900	38.7	367	4 AAB30279	Aab30279 Peptide #
19	1900	38.7	367	4 AAM56268	Aam56268 Human bra
20	1900	38.7	367	4 AAM04186	Aam04186 Peptide #
21	1900	38.7	367	5 ABR38824	Abv38824 Human pep
22	1286	26.2	242	7 ADD26607	Add26607 Human adi
23	923	18.8	128	6 ABt11844	Abt11844 Human MDD
24	503	10.3	113	4 ABR61382	Abt61382 Partial r
25	503	10.3	113	7 ADD27822	Add27822 Rat alpha

26	503	10.3	113	7 ABR01494	Abv01494 Partial r
27	503	10.3	113	7 ABR01543	Abv01543 Partial r
28	488	9.9	652	4 ABR59094	Abv59094 Drosophila
29	479	9.8	502	4 ABR61383	Abt61383 Partial D
30	479	9.8	502	7 ADD27824	Add27824 Fruit fly
31	479	9.8	502	7 ABR01495	Abv01495 Partial f
32	479	9.8	502	7 ABR01544	Abv01544 Partial f
33	349	7.1	68	4 ABR38187	Abb38187 Peptide #
34	349	7.1	68	4 AAM31616	Aam31616 Peptide #
35	349	7.1	68	4 AAM71337	Aam71337 Human bon
36	349	7.1	68	4 AAM58821	Aam58821 Human bra
37	349	7.1	68	4 ABR53045	Abv53045 Human liv
38	349	7.1	68	5 ABR41140	Abv41140 Human pep
39	265	5.4	545	2 AAM79296	Aam79296 Neisseria
40	255.5	5.2	546	7 ABR74499	Abv74499 Mycobacte
41	251	5.1	63	6 ABR63220	Abv63220 Clone T11
42	251	5.1	63	6 ABR43797	Abv43797 Clone T11
43	240	4.9	364	6 ADA34095	Ada34095 Actinotoba
44	173.5	3.5	2469	5 AAE18207	Aae18207 Human MOL
45	173.5	3.5	2469	7 ADD18192	Add18192 Human mol

ALIGNMENTS

RESULT 1
ABR61376
ID ABR61376 standard; protein; 928 AA.
XX ABR61376;
AC
XX
XX 01-AUG-2003 (first entry)
DT
XX
XX Human GlcNAc-phosphotransferase alpha-subunit.
DE
XX
XX Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;
KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;
KW lysosomal storage disease; enzyme; alpha-subunit.
XX
XX Homo sapiens.
OS
XX
XX
XX US6537785-B1.
PN
XX
XX 25-MAR-2003.
PD
XX
XX 10-AUG-2000; 2000US-00636077.
PF
XX
XX 14-SEP-1999; 99US-0153831P.
PR
XX
XX (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.
PA
XX
XX Canfield WM;
XX
XX NPI; 2001-290356/30.
DR
XX
XX N-PSDB; ACC61001.
PT
XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.
PT
XX
XX Disclosure; Page 26-28; 62pp; English.
PS
XX
XX The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has nephrotropic activity, and may be useful in enzyme replacement therapy. A protein of the invention (I), (II) is useful for preparing a phosphorylated lysosomal hydrolase. The phosphorylated hydrolase comprising a terminal mannose-6-phosphate, is useful for treating a

CC patient suffering from a lysosomal storage disease. The present sequence
 CC is used in the exemplification of the invention

XX Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 4; Length 928;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MFKLQRTYTCLSHRVGLVYCFGLVVTIVSAQFGVVLWMSRDQYHVLFDSDYRDI 60
DB 1 MFKLQRTYTCLSHRVGLVYCFGLVVTIVSAQFGVVLWMSRDQYHVLFDSDYRDI 60
QY 61 AGSFQNRCLPMPIDVYTWNGTDLLELKLQVRBOMEBOQAMEIIGKNTTEPTK 120
DB 61 AGSFQNRCLPMPIDVYTWNGTDLLELKLQVRBOMEBOQAMEIIGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKVPMLVLDPALPANTTLKDVPSLYSPHSASDIFNVAKPKPSTNV 180
DB 121 KSEKQLECLTHCIKVPMLVLDPALPANTTLKDVPSLYSPHSASDIFNVAKPKPSTNV 180
QY 181 SYVVPSTKDVEDAHSGLLKGNRSQTVWRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
DB 181 SYVVPSTKDVEDAHSGLLKGNRSQTVWRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
QY 181 SYVVPSTKDVEDAHSGLLKGNRSQTVWRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
DB 181 SYVVPSTKDVEDAHSGLLKGNRSQTVWRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
QY 241 NQKTKLPENLSKVKYLQLYSEASVALIKLNPKDFOLNKQTKNMTIDGKELTISPA 300
DB 241 NQKTKLPENLSKVKYLQLYSEASVALIKLNPKDFOLNKQTKNMTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKQODEDISASRFEDNEELRYSLSIERHAPVNRNIFVTNGQIPSMNL 360
DB 301 YLLMDLSAISQSKQODEDISASRFEDNEELRYSLSIERHAPVNRNIFVTNGQIPSMNL 360
QY 361 DNPRTVITHQDFENLSHLPTFSSPAIESHIRELEGSOKRIYLNDDVMGKDWPPDF 420
DB 361 DNPRTVITHQDFENLSHLPTFSSPAIESHIRELEGSOKRIYLNDDVMGKDWPPDF 420
QY 421 YHSHSQQKYLLTPWPNCAEGCPGSMYKGYCDXACNNACDWDGDCSGNSGGRYIAG 480
DB 421 YHSHSQQKYLLTPWPNCAEGCPGSMYKGYCDXACNNACDWDGDCSGNSGGRYIAG 480
QY 481 GGGTGSIGVGHFMQFGGINSYSYCNQGCANSMWADKFCDDQCNVLSGFPADGQGDHF 540
DB 481 GGGTGSIGVGHFMQFGGINSYSYCNQGCANSMWADKFCDDQCNVLSGFPADGQGDHF 540
QY 541 HELYKVILLPNQTHYIIPKGECLPYSPFAEVAKRGVEGAYSDNPIIRHASIANWKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYSPFAEVAKRGVEGAYSDNPIIRHASIANWKTIHL 600
QY 601 IMHSGNATTHIHNLTQNTNDEEFKMOITVEVDTRBGPXNSTQKGYENLVSPITLLP 660
DB 601 IMHSGNATTHIHNLTQNTNDEEFKMOITVEVDTRBGPXNSTQKGYENLVSPITLLP 660
QY 661 EABILFEDIPEKERPFKRDVNSTRAQEBVKIPLVNISLPPDAQLSINTDLQLEH 720
DB 661 EABILFEDIPEKERPFKRDVNSTRAQEBVKIPLVNISLPPDAQLSINTDLQLEH 720
QY 721 GGIITLKGYNLSSALLRSFLMNSQAKTKNQAIITDENRDSIVAQEQVHKSILPNSLG 780
DB 721 GGIITLKGYNLSSALLRSFLMNSQAKTKNQAIITDENRDSIVAQEQVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKNGDQGNPDLLETTARFVETHTQKTIGNVKEKPPSLIV 840
DB 781 VSERLQRLTFPAVSVKNGDQGNPDLLETTARFVETHTQKTIGNVKEKPPSLIV 840
QY 841 PLESQWTKKKITGKEKENSMEENAEHNIQTEVLLGRKIQHTYDSYLGFLPWEKKKYF 900
DB 841 PLESQWTKKKITGKEKENSMEENAEHNIQTEVLLGRKIQHTYDSYLGFLPWEKKKYF 900
QY 901 QDLDEESLKTOLAFYTSKNTGRQLK 928
DB 901 QDLDEESLKTOLAFYTSKNTGRQLK 928
QY 901 QDLDEESLKTOLAFYTSKNTGRQLK 928
DB 901 QDLDEESLKTOLAFYTSKNTGRQLK 928

```

RESULT 2

ID ADD27812 standard; protein; 928 AA.

AC ADD27812;

DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase alpha subunit.

KM human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
 KM UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.

OS Homo sapiens.

PN US2003119088-A1.

PD 26-JUN-2003.

PF 21-DEC-2001; 2001US-00023888.

PR 21-DEC-2001; 2001US-00023888.

PA (NOVA-) NOVAZYME PHARM INC.

PI Canfield W, Kudo M;

DR WPI; 2003-801323/75.

DR N-PSDB; ADD27811.

PT Phosphorylating a protein for treating a patient suffering from a
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
 PT protein.

PS Claim 7; SEQ ID NO 4; 55pp; English.

XX The invention relates to a method of phosphorylating a protein comprising
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
 CC acetylglucosamine) and producing a phosphorylated protein. The method is
 CC useful for treating a patient suffering from a lysosomal storage disease
 CC e.g. Fabry's disease. The present sequence represents the amino acid
 CC sequence of the human GlcNAc-phosphotransferase alpha subunit.

XX Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; Length 928;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MFKLQRTYTCLSHRVGLVYCFGLVVTIVSAQFGVVLWMSRDQYHVLFDSDYRDI 60
DB 1 MFKLQRTYTCLSHRVGLVYCFGLVVTIVSAQFGVVLWMSRDQYHVLFDSDYRDI 60
QY 61 AGSFQNRCLPMPIDVYTWNGTDLLELKLQVRBOMEBOQAMEIIGKNTTEPTK 120
DB 61 AGSFQNRCLPMPIDVYTWNGTDLLELKLQVRBOMEBOQAMEIIGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKVPMLVLDPALPANTTLKDVPSLYSPHSASDIFNVAKPKPSTNV 180
DB 121 KSEKQLECLTHCIKVPMLVLDPALPANTTLKDVPSLYSPHSASDIFNVAKPKPSTNV 180
QY 181 SYVVPSTKDVEDAHSGLLKGNRSQTVWRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
DB 181 SYVVPSTKDVEDAHSGLLKGNRSQTVWRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
QY 181 SYVVPSTKDVEDAHSGLLKGNRSQTVWRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
DB 181 SYVVPSTKDVEDAHSGLLKGNRSQTVWRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
QY 241 NQKTKLPENLSKVKYLQLYSEASVALIKLNPKDFOLNKQTKNMTIDGKELTISPA 300
DB 241 NQKTKLPENLSKVKYLQLYSEASVALIKLNPKDFOLNKQTKNMTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKQODEDISASRFEDNEELRYSLSIERHAPVNRNIFVTNGQIPSMNL 360
DB 301 YLLMDLSAISQSKQODEDISASRFEDNEELRYSLSIERHAPVNRNIFVTNGQIPSMNL 360

```

```
Db 301 YLIMDLAISQSKODEDISASRFEDNEELRYSLSRIERHAPWVRNIFIVTNGQIPSWMLN 360
Qy 361 DNRVTVITHQDVFNRLSHLPTFSSPAIESHRIEGLSQKFIYNDVWFGKDVWPDDE 420
Db 361 DNRVTVITHQDVFNRLSHLPTFSSPAIESHRIEGLSQKFIYNDVWFGKDVWPDDE 420
Qy 421 YSHSKQKQVYLTWPVNCABGCGSWIKDGYCDKACNNSACMDGDCGSGSGSRYYAG 480
Db 421 YSHSKQKQVYLTWPVNCABGCGSWIKDGYCDKACNNSACMDGDCGSGSGSRYYAG 480
Qy 481 GGGTGSIGVGHMPQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCFDAGDCGDHF 540
Db 481 GGGTGSIGVGHMPQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCFDAGDCGDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYPSFAVAKRGVEGAYSDNPIIRHNASINKKKTITLH 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYPSFAVAKRGVEGAYSDNPIIRHNASINKKKTITLH 600
Qy 601 IMHSGMNAATTIHFNLTFOUNTDEEFMQITVEVDTRREGPKLNTAQKGYENLVSPITLLP 660
Db 601 IMHSGMNAATTIHFNLTFOUNTDEEFMQITVEVDTRREGPKLNTAQKGYENLVSPITLLP 660
Qy 661 EAEILFEDIPEKRFPEKFRHDVNSTRAQEVEKIPLVNISSLPRDAQSLNTLDLQLEH 720
Db 661 EAEILFEDIPEKRFPEKFRHDVNSTRAQEVEKIPLVNISSLPRDAQSLNTLDLQLEH 720
Qy 721 GDTTLKGYNLKSALIRSFVLSNQAIIITDETNDSLVADQEKVHKSIIPLNSLG 780
Db 721 GDTTLKGYNLKSALIRSFVLSNQAIIITDETNDSLVADQEKVHKSIIPLNSLG 780
Qy 781 VSERLQRLTFPAVSVKNGHDOGNPDLLETTAFRVEYTHQKTIIGNVTEKRPPLIIV 840
Db 781 VSERLQRLTFPAVSVKNGHDOGNPDLLETTAFRVEYTHQKTIIGNVTEKRPPLIIV 840
Qy 841 PLESQVTKKKITGKEKENSMEENAEENHIGYTEVLGRKLOHYTDSYLGFLPWEKKYCF 900
Db 841 PLESQVTKKKITGKEKENSMEENAEENHIGYTEVLGRKLOHYTDSYLGFLPWEKKYCF 900
Qy 901 ODLDDEESLKTQLAIFYTDSKNTGRQLK 928
Db 901 ODLDDEESLKTQLAIFYTDSKNTGRQLK 928

RESULT 3
ABW01488
ID ABW01488 standard; protein; 928 AA.
XX
AC ABW01488;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GlcNAc-phosphotransferase alpha subunit precursor protein.
XX
KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
KW galectin; human; enzyme; lectin resistant cell;
KW deoxymannojirimycin; kifunensine; glycosylation inhibition.
XX
OS Homo sapiens.
XX
PN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023889.
XX
PR 21-DEC-2001; 2001US-00023889.
XX
PA (NOVA-) NOVARTIS PHARM INC.
XX
PI Canfield WM;
XX
DR WPI; 2003-810984/76.
```

```
DR N-PSDB; AAD62491.
XX
PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
PS Claim 10; Page 15-18; 46pp; English.
XX
CC The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is human N-
CC acetylglucosamine-1 (GlcNAc)-phosphotransferase alpha subunit precursor
CC protein
SQ Sequence 928 AA:
Query Match 100.0%; Score 4907; DB 7; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLFELLQROTYYTCLSHRYGLVCFGLGVVTVIVSAFQGEVVLBMSRDQYVHLPDSYRDNI 60
Db 1 MLFELLQROTYYTCLSHRYGLVCFGLGVVTVIVSAFQGEVVLBMSRDQYVHLPDSYRDNI 60
Qy 61 AGKSFQNRRLCPMPIDVYTVWNGTDLBELIKLQOVRBQEBEQKAMREILGNTEPTK 120
Db 61 AGKSFQNRRLCPMPIDVYTVWNGTDLBELIKLQOVRBQEBEQKAMREILGNTEPTK 120
Qy 121 KSEKQJECCLTHGCKVPMVLVDLPALPANITLTKQVSPSYPSFHSASQIFNVAKPNSTVY 180
Db 121 KSEKQJECCLTHGCKVPMVLVDLPALPANITLTKQVSPSYPSFHSASQIFNVAKPNSTVY 180
Qy 121 KSEKQJECCLTHGCKVPMVLVDLPALPANITLTKQVSPSYPSFHSASQIFNVAKPNSTVY 180
Db 121 KSEKQJECCLTHGCKVPMVLVDLPALPANITLTKQVSPSYPSFHSASQIFNVAKPNSTVY 180
Qy 181 SVVPFDSTKQVEDAHSGLLKNGSRQTVWRGYLTITDXEVPGLVIMQDLAFLSGFPPTKET 240
Db 181 SVVPFDSTKQVEDAHSGLLKNGSRQTVWRGYLTITDXEVPGLVIMQDLAFLSGFPPTKET 240
Qy 241 NQKTKLPENLSSKVKLQIYSEASVALLKLNPKDPOELNQTCKNMITDGEILTISA 300
Db 241 NQKTKLPENLSSKVKLQIYSEASVALLKLNPKDPOELNQTCKNMITDGEILTISA 300
Qy 301 YLIMDLAISQSKODEDISASRFEDNEELRYSLSRIERHAPWVRNIFIVTNGQIPSWMLN 360
Db 301 YLIMDLAISQSKODEDISASRFEDNEELRYSLSRIERHAPWVRNIFIVTNGQIPSWMLN 360
Qy 361 DNRVTVITHQDVFNRLSHLPTFSSPAIESHRIEGLSQKFIYNDVWFGKDVWPDDE 420
Db 361 DNRVTVITHQDVFNRLSHLPTFSSPAIESHRIEGLSQKFIYNDVWFGKDVWPDDE 420
Qy 421 YSHSKQKQVYLTWPVNCABGCGSWIKDGYCDKACNNSACMDGDCGSGSGSRYYAG 480
Db 421 YSHSKQKQVYLTWPVNCABGCGSWIKDGYCDKACNNSACMDGDCGSGSGSRYYAG 480
Qy 481 GGGTGSIGVGHMPQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCFDAGDCGDHF 540
Db 481 GGGTGSIGVGHMPQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCFDAGDCGDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYPSFAVAKRGVEGAYSDNPIIRHNASINKKKTITLH 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYPSFAVAKRGVEGAYSDNPIIRHNASINKKKTITLH 600
Qy 601 IMHSGMNAATTIHFNLTFOUNTDEEFMQITVEVDTRREGPKLNTAQKGYENLVSPITLLP 660
Db 601 IMHSGMNAATTIHFNLTFOUNTDEEFMQITVEVDTRREGPKLNTAQKGYENLVSPITLLP 660
Qy 661 EAEILFEDIPEKRFPEKFRHDVNSTRAQEVEKIPLVNISSLPRDAQSLNTLDLQLEH 720
```

Dh 661 EAELIFEDIPKPKRPPKFRHDVNSTRAQEEVKIPLVNIISLLPDAQLSINTLDLQLEH 720
Qy 721 GDTTLKGYNLKSALRSFLMNSQAKIKNOAIITDETNDLSVAPQEQVHKSIIPLNSLG 780
Db 721 GDTTLKGYNLKSALRSFLMNSQAKIKNOAIITDETNDLSVAPQEQVHKSIIPLNSLG 780
Qy 781 VSERLQRLTFPAVSAYKNGHDGQNPDLLETTARFRVETHQKTIIGAVTKEKPPSLIV 840
Db 781 VSERLQRLTFPAVSAYKNGHDGQNPDLLETTARFRVETHQKTIIGAVTKEKPPSLIV 840
Qy 841 PLESQMTKEKKITGKEKENSMEENAHNIGTVEVLGRKLOHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQMTKEKKITGKEKENSMEENAHNIGTVEVLGRKLOHYTDSYLGFLPWEKKKYF 900
Qy 901 ODLLDEESLKTOLAYFTDSKNTGRQLK 928
Db 901 ODLLDEESLKTOLAYFTDSKNTGRQLK 928

RESULT 4

ABW01537 standard; protein; 928 AA.
ID ABW01537
XX
AC ABW01537;
XX
DT 15-JUN-2004 (first entry)
XX
DE Human GlcNAc-phosphotransferase alpha subunit precursor protein.
XX
KM Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
KW N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.
XX
OS Homo sapiens.
XX
PN US2003124653-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023890.
XX
PR 21-DEC-2001; 2001US-00023890.
XX
PA (NOVA-) NOVAZYME PHARM INC.
XX
PI Canfield WM;
XX
DR WPI, 2003-810985/76.
XX
DR N-PDB; AAD62650.
XX
PT Producing a glycoprotein with reduced complex carbohydrates by culturing
PT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.
XX
PS Claim 10; Page 16-18; 46pp; English.
XX
CC The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing the lectin resistant
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-
CC phosphotransferase alpha subunit precursor protein
XX
SQ Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; Length 928;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MFKLLQROTYTCLSHRYGLYVCFGLVAVTTIVSAFQFGSEVLEMSRDQYHVLFDSDYDNI 60
Db 1 MFKLLQROTYTCLSHRYGLYVCFGLVAVTTIVSAFQFGSEVLEMSRDQYHVLFDSDYDNI 60

Qy 61 AGSFQNRCLPMPIDVYTWVNGTDLLELLELQVREOMEEQKAMREILCKNTTEPTK 120
Db 61 AGSFQNRCLPMPIDVYTWVNGTDLLELLELQVREOMEEQKAMREILCKNTTEPTK 120
Qy 121 KSEKQECILTHQIKYPMILVDPALPANITLTDVSLYSFHSASDIFVAVKPKPSTNV 180
Db 121 KSEKQECILTHQIKYPMILVDPALPANITLTDVSLYSFHSASDIFVAVKPKPSTNV 180
Qy 181 SVVPFDSTKDVEDAHSGLLKGNRSQRTWAGYLTDXEVLVIMODLAFSLGFPPTFKET 240
Db 181 SVVPFDSTKDVEDAHSGLLKGNRSQRTWAGYLTDXEVLVIMODLAFSLGFPPTFKET 240
Qy 241 NQKTLPLPNSKYVLLDLYGBASVALLKLNPNPDQCLNKOTKKNMTIDKELITISA 300
Db 241 NQKTLPLPNSKYVLLDLYGBASVALLKLNPNPDQCLNKOTKKNMTIDKELITISA 300
Qy 301 YLLMDLSAISQSKODEDISASRPEDNEELRYSLSIERHAPVWRNFIYVNGQIPSWML 360
Db 301 YLLMDLSAISQSKODEDISASRPEDNEELRYSLSIERHAPVWRNFIYVNGQIPSWML 360
Qy 361 DNEPRTIIVTHQVFNRLSHLPTSSPALSHIRIEGLSQKFIYLANDVMFGKVPDPDF 420
Db 361 DNEPRTIIVTHQVFNRLSHLPTSSPALSHIRIEGLSQKFIYLANDVMFGKVPDPDF 420
Qy 421 YSHSKQKXYLTWPVNCAGCGPSWIKDGYCDKACNNSACDMDGDCSGNSGSRYYIAG 480
Db 421 YSHSKQKXYLTWPVNCAGCGPSWIKDGYCDKACNNSACDMDGDCSGNSGSRYYIAG 480
Qy 481 GGGTGSIGVGHWPQFGGINSVAYCNCAGNSWLADKFCDOACNVLSGCFDAGCGQDHF 540
Db 481 GGGTGSIGVGHWPQFGGINSVAYCNCAGNSWLADKFCDOACNVLSGCFDAGCGQDHF 540
Qy 541 HELYKYLILPNQTHYIIPKGECLPYSPFAVEAKRGVGAYSNPPIIRHASIANKMTIHL 600
Db 541 HELYKYLILPNQTHYIIPKGECLPYSPFAVEAKRGVGAYSNPPIIRHASIANKMTIHL 600
Qy 601 IMHSGNATTIHNLTFQNTNDEEFKQITVEVDTRBQKLNSTAKQGYENLVSPITLLP 660
Db 601 IMHSGNATTIHNLTFQNTNDEEFKQITVEVDTRBQKLNSTAKQGYENLVSPITLLP 660
Qy 661 EAELIFEDIPKPKRPPKFRHDVNSTRAQEEVKIPLVNIISLLPDAQLSINTLDLQLEH 720
Db 661 EAELIFEDIPKPKRPPKFRHDVNSTRAQEEVKIPLVNIISLLPDAQLSINTLDLQLEH 720
Qy 721 GDTTLKGYNLKSALRSFLMNSQAKIKNOAIITDETNDLSVAPQEQVHKSIIPLNSLG 780
Db 721 GDTTLKGYNLKSALRSFLMNSQAKIKNOAIITDETNDLSVAPQEQVHKSIIPLNSLG 780
Qy 781 VSERLQRLTFPAVSAYKNGHDGQNPDLLETTARFRVETHQKTIIGAVTKEKPPSLIV 840
Db 781 VSERLQRLTFPAVSAYKNGHDGQNPDLLETTARFRVETHQKTIIGAVTKEKPPSLIV 840
Qy 841 PLESQMTKEKKITGKEKENSMEENAHNIGTVEVLGRKLOHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQMTKEKKITGKEKENSMEENAHNIGTVEVLGRKLOHYTDSYLGFLPWEKKKYF 900
Qy 901 ODLLDEESLKTOLAYFTDSKNTGRQLK 928
Db 901 ODLLDEESLKTOLAYFTDSKNTGRQLK 928

RESULT 5

AAE25290 standard; protein; 1256 AA.

AAE25290;

AC AAE25290;

DT 30-OCT-2002 (first entry)

Qy Human nucleic acid-associated protein (NAAP-9).
Db Human nucleic acid-associated protein; NAAP-9; neurological disorder;
KW arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;

KM Lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
KM autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
KM gene therapy; neurotropic; neuroprotective; cerebroprotective; virucide;
KM immunosuppressive; protozoacide; antimicrobial.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 448..469
FT /note="Notch domain"
FT Domain 500..536
FT /note="Notch domain"
FT Domain 1018..1030
FT /note="EF-hand calcium-binding domain"
PN W0200250279-A2.
XX
PD 27-JUN-2002.
XX
PF 19-DEC-2001; 2001WO-US050256.
XX
PR 21-DEC-2000; 2000US-0257714P.
XX
PR 05-JAN-2001; 2001US-0260081P.
PR 16-JAN-2001; 2001US-0262302P.
PR 23-JAN-2001; 2001US-0263823P.
PR 02-FEB-2001; 2001US-0266088P.
PR 29-OCT-2001; 2001US-0348442P.
XX
XX
PA (INCYTE GENOMICS INC.
XX
PI Baughin MR, Lu Y, Arvizu C, Rankumar J, Yao MG, Policky JL,
PI Walla NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG,
PI Borowsky JM, Lu DM, Gandhi AR, Griffin JA, Xu Y, Azimtai Y,
PI Glezen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;
PI Lee EA, Yang J, Gorvad AE, Emerling BM, Marquis CP, Lee SY;
PI Swarnakar A, Reddy R;
XX
DR WPI; 2002-519887/55.
DR N-PSDB; AAD41199.
XX
XX Nucleic acid associated proteins and nucleic acids for diagnosing,
PT treating and preventing cell proliferative (e.g. cancers), neurological
PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).
XX
XX Claim 64; Page 163-165; 193pp; English.
XX
XX The invention relates to nucleic acid-associated proteins (NAAP) and
CC nucleic acids. The nucleic acid and amino acid sequences are useful for
CC diagnosing, treating and preventing cell proliferative e.g.
CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological
CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections
CC (e.g. malaria, or leishmania), as well as in assessing the effects of
CC exogenous compound on the expression of nucleic acid and amino acid
CC sequences of nucleic acid-associated proteins. The invention is useful in
CC gene therapy. The present sequence is human NAAP-9
XX
SQ Sequence 1256 AA;
XX
Query Match 99.8%; Score 4896; DB 5; Length 1256;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 926; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MFKLLQRTYCTLSHRVLYCYCFGVVVTISAFQFSEVVLWMSRDQYHLFDSYRNI 60
XX
DB 1 MFKLLQRTYCTLSHRVLYCYCFGVVVTISAFQFSEVVLWMSRDQYHLFDSYRNI 60
QY 61 ACKSFQNRCLPMPIDVYVTVNGTDLLELKLQVREOMEBEQAKMEIIGKNTTEPTK 120
XX
DB 61 ACKSFQNRCLPMPIDVYVTVNGTDLLELKLQVREOMEBEQAKMEIIGKNTTEPTK 120
QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDLPBLYPSFHSASDIPNVAKPKNSTNV 180
XX
121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDLPBLYPSFHSASDIPNVAKPKNSTNV 180

DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDLPBLYPSFHSASDIPNVAKPKNSTNV 180
QY 181 SVVVFSDTKQVEDAHSGLLKGNRSQYVWRGYLTDDXVPEGLVIMQDLAFSGPPPFKET 240
DB 181 SVVVFSDTKQVEDAHSGLLKGNRSQYVWRGYLTDDXVPEGLVIMQDLAFSGPPPFKET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALIKLANPKDPOEANKTKNMITDGKELITSPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALIKLANPKDPOEANKTKNMITDGKELITSPA 300
QY 301 YLLMDLSAISQSKODEDISASREDENEELRYSLSIRERAPVVRNIFIVTNGQIPSMNL 360
DB 301 YLLMDLSAISQSKODEDISASREDENEELRYSLSIRERAPVVRNIFIVTNGQIPSMNL 360
QY 361 DNPRTVTVHQDYFRNLSHLPTSSPAISHIRIEGLSQKFTYLNDDVWFGKDWPPDE 420
DB 361 DNPRTVTVHQDYFRNLSHLPTSSPAISHIRIEGLSQKFTYLNDDVWFGKDWPPDE 420
QY 421 YSHSKGQKYLTPVPNCAEGCPSWI KDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480
DB 421 YSHSKGQKYLTPVPNCAEGCPSWI KDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480
QY 481 GGGTGSTGVGHFWQFGGGLNSVYCNQGCANSMYADKFCDOACNVLSGCFDAGDCQDHF 540
DB 481 GGGTGSTGVGHFWQFGGGLNSVYCNQGCANSMYADKFCDOACNVLSGCFDAGDCQDHF 540
QY 541 HELYKVILPNOYHYIIPKGECLPYFSPAFAVARGVAGASDPIIRHASIANKMTIHL 600
DB 541 HELYKVILPNOYHYIIPKGECLPYFSPAFAVARGVAGASDPIIRHASIANKMTIHL 600
QY 601 IMHSGNATTHFNLTFOINTDEEFKQITVEVTEEGPKLNTSTAQGYENLVSPITLLP 660
DB 601 IMHSGNATTHFNLTFOINTDEEFKQITVEVTEEGPKLNTSTAQGYENLVSPITLLP 660
QY 661 EARLPEIDPKERPFKPKRHDVNSRBAOEYKIRLVNLSILPKDAQLSINTLDLQLEH 720
DB 661 EARLPEIDPKERPFKPKRHDVNSRBAOEYKIRLVNLSILPKDAQLSINTLDLQLEH 720
QY 721 GDITLKGYNLSKSAALLRFLMNSQAHAKIRNOAITDETNDLSVAPQEKVHKSILPNSIG 780
DB 721 GDITLKGYNLSKSAALLRFLMNSQAHAKIRNOAITDETNDLSVAPQEKVHKSILPNSIG 780
QY 781 VSRRLQRLTPPAVSVKVNGDQGNPDLLETTARFRVETHQKTIGSVTKKPSLIY 840
DB 781 VSRRLQRLTPPAVSVKVNGDQGNPDLLETTARFRVETHQKTIGSVTKKPSLIY 840
QY 841 PLSQMTKEKKIRGKEKENSREMEANHIGTEVLLGRKLOHYTDSYIGFLPWEKKKYF 900
DB 841 PLSQMTKEKKIRGKEKENSREMEANHIGTEVLLGRKLOHYTDSYIGFLPWEKKKYF 900
QY 901 QDILDEBSIKTQLAVFTDSKNTGRQLK 928
DB 901 QDILDEBSIKTQLAVFTDSKNTGRQLK 928
XX
RESULT 6
ADD27810
ID ADD27810 standard; protein; 1199 AA.
XX
AC ADD27810;
XX
DT 15-JUN-2004 (first entry)
XX
DE Soluble human GlcNAc-phosphotransferase.
XX
XX human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
KM UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease;
XX enzyme.
OS Synthetic.
XX Homo sapiens.
XX FN US2003119088-A1.

```

XX 26-JUN-2003.
PD 21-DEC-2001; 2001US-00023888.
PF 21-DEC-2001; 2001US-00023888.
PR 21-DEC-2001; 2001US-00023888.
XX (NOVA-) NOVAAZYME PHARM INC.
XX PA
XX PI Canfield W, Kudo M;
XX DR WPI; 2003-801323/75.
XX DR N-PSDB; ADD27809.
XX PT Phosphorylating a protein for treating a patient suffering from a
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
PT protein.
XX PS
XX PS Claim 3; SEQ ID NO 2; 55pp; English.
XX CC The invention relates to a method of phosphorylating a protein comprising
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of soluble human GlcNAc-phosphotransferase.
XX CC
XX Sequence 1199 AA:
Query Match 94.4%; Score 4630; DB 7; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 45 SSDQYVLPDSRDNDAGSKFONRLCLPMDIDVYTYWNGTDLLEKELQVREOMESEQ 104
DB 35 SSDQYVLPDSRDNDAGSKFONRLCLPMDIDVYTYWNGTDLLEKELQVREOMESEQ 94
QY 105 KAMREILGNKTEPTKSEKOLECLTHCIRKVMVLDPALPANITLKDVPSELYSEFSA 164
DB 95 KAMREILGNKTEPTKSEKOLECLTHCIRKVMVLDPALPANITLKDVPSELYSEFSA 154
QY 165 SDIFNVAKKPNSTNVSVVFPDSTKDVEDAHSGLLKNGSRQVWVGYYLTTRKEVGLVLM 224
DB 155 SDIFNVAKKPNSTNVSVVFPDSTKDVEDAHSGLLKNGSRQVWVGYYLTTRKEVGLVLM 214
QY 225 QDLAPLGGPPTPEKTNQLKTKLPENLSKXVLLQLYSBASVALLKLNPPKDFOLNKQT 284
DB 215 QDLAPLGGPPTPEKTNQLKTKLPENLSKXVLLQLYSBASVALLKLNPPKDFOLNKQT 274
QY 285 KKNMTIDGKELTISPAVLLMDLSAISQSKODEDISASRPEDNEELRYSLRSTERRAPVVR 344
DB 275 KKNMTIDGKELTISPAVLLMDLSAISQSKODEDISASRPEDNEELRYSLRSTERRAPVVR 334
QY 345 NIFIVTNGQIPSMVLNDNPRVTIVTHQDFRNLISHLPTFFSSPAISHIRIEGLSQKFIY 404
DB 335 NIFIVTNGQIPSMVLNDNPRVTIVTHQDFRNLISHLPTFFSSPAISHIRIEGLSQKFIY 394
QY 405 LNDVWVFGQVWPDFFYSKSGQKYLTPVPCNACBGGPGSVTKGICDCAKNNASACMD 464
DB 395 LNDVWVFGQVWPDFFYSKSGQKYLTPVPCNACBGGPGSVTKGICDCAKNNASACMD 454
QY 465 GDDCGSGSGSRYIAGGGGTGSIQVGHQPGQGGINSVSYCNGGANSMLDKDFDQACN 524
DB 455 GDDCGSGSGSRYIAGGGGTGSIQVGHQPGQGGINSVSYCNGGANSMLDKDFDQACN 514
QY 525 VLSGCFDAGDCGQDFHEHLYKVIILPNQTHYIIPYGECLPYFSFAEVAKGVGAYSNP 584
DB 515 VLSGCFDAGDCGQDFHEHLYKVIILPNQTHYIIPYGECLPYFSFAEVAKGVGAYSNP 574
QY 585 IIRHASIAKMKTIHLIMHSGNNAATTIHNLTFQUNDEEFKQITVEVDREGFKNST 644
DB 575 IIRHASIAKMKTIHLIMHSGNNAATTIHNLTFQUNDEEFKQITVEVDREGFKNST 634

```

```

QY 645 AOKGYENLVSPITLPEAEILFEDIPKKEKPKKHVDVNSRRQAEVKIPLVNIISLP 704
DB 635 AOKGYENLVSPITLPEAEILFEDIPKKEKPKKHVDVNSRRQAEVKIPLVNIISLP 694
QY 705 KDAQLSLNTLDQLEHGDITLKGYNLSKSLRSPFLMNSQAKIKNOAIITDETDSLVA 764
DB 695 KDAQLSLNTLDQLEHGDITLKGYNLSKSLRSPFLMNSQAKIKNOAIITDETDSLVA 754
QY 765 PQEKQYHKSILPNSLGVSERLQRLTPPAVSVCVNGHDQGNPDLLETTARFRVETHQK 824
DB 755 PQEKQYHKSILPNSLGVSERLQRLTPPAVSVCVNGHDQGNPDLLETTARFRVETHQK 814
QY 825 TIGGANTKEKPSLIVPLESOMTKKIKYGEKENSMEENAEHIGTEVLLGRLQHY 884
DB 815 TIGGANTKEKPSLIVPLESOMTKKIKYGEKENSMEENAEHIGTEVLLGRLQHY 874
QY 885 TDSYLGLPWEKKKRYQDILLDEBSLKTOLAYFTTSKATGR 925
DB 875 TDSYLGLPWEKKKRYQDILLDEBSLKTOLAYFTTSKATGR 915

RESULT 7
ABM01487
ID ABM01487 standard; protein; 1199 AA.
AC ABM01487;
XX AC
XX DT 15-JAN-2004 (first entry)
XX DE
XX DE N-acetylglucosamine-1(GlcNAc)-phosphotransferase.
XX KM Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
XX KM lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
XX KM gastrointestinal; enzyme; lectin resistant cell; deoxymannojirimycin;
XX KM kifunensine; glycosylation inhibition.
XX OS
XX OS Unidentified.
XX PN US2003124652-A1.
XX PD 03-JUL-2003.
XX PE 21-DEC-2001; 2001US-00023889.
XX PR 21-DEC-2001; 2001US-00023889.
XX PA (NOVA-) NOVAAZYME PHARM INC.
XX PI Canfield WM;
XX DR WPI; 2003-810984/76.
XX DR N-PSDB; ADD2490.
XX PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprising culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX PS
XX PS Claim 8; Page 10-13; 46pp; English.
XX CC The invention relates to a method for producing a high mannose
XX CC glycoprotein. The method comprises: introducing and expressing a
XX CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
XX CC the cell in the presence of a lectin to obtain a lectin resistant cell;
XX CC isolating the cell; culturing the cell in the presence of
XX CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
XX CC glycoprotein; and collecting the glycoprotein. The invention is useful in
XX CC gene therapy. The method is useful for producing a high mannose
XX CC glycoprotein in a complex carbohydrate deficient cell for treating
XX CC lysosomal storage disease. The present sequence is N-acetylglucosamine-1
XX CC (GlcNAc)-phosphotransferase
XX Sequence 1199 AA:

```

Query Match		94.4%;	Score 4630;	DB 7;	Length 1199;
Best Local Similarity		99.3%;	Pred. No. 0;		
Matches 875;		Conservative 2;	Mismatches 4;	Indels 0;	Gaps 0;
QY	45	SSDOYHVLFDPSYRDNIAKGSFQNRCLPMPIDVYTWVNGTDLELKELOQVREQMEEQ	104		
DB	35	SSDOYHVLFDPSYRDNIAKGSFQNRCLPMPIDVYTWVNGTDLELKELOQVREQMEEQ	94		
QY	105	KAMREILGKNTTEPTKSEKQLECLTHCIKVMVLDPALPANITLKDVPSLPSFSA	164		
DB	95	KAMREILGKNTTEPTKSEKQLECLTHCIKVMVLDPALPANITLKDVPSLPSFSA	154		
QY	165	SDIFNVAKPKNSTVSVVFPDSTKDVEDAHSGLLKGNRSQVWVRGYLTDEVPGLVLM	224		
DB	155	SDIFNVAKPKNSTVSVVFPDSTKDVEDAHSGLLKGNRSQVWVRGYLTDEVPGLVLM	214		
QY	225	QDLAFLSGPPTFKETNOUKTYLPENLSKVKLLQLYSEASVALIKLNPKDFQELNKOT	284		
DB	215	QDLAFLSGPPTFKETNOUKTYLPENLSKVKLLQLYSEASVALIKLNPKDFQELNKOT	274		
QY	285	KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEELRYSLSIRHAPWVR	344		
DB	275	KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEELRYSLSIRHAPWVR	334		
QY	345	NIPIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIREGLSQKFIY	404		
DB	335	NIPIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIREGLSQKFIY	394		
QY	405	LNDDVWFGKDVWPDDEFYSHSKQKYLLTPVPNCABGCGPSWIKDGYCDKACNNSACMD	464		
DB	395	LNDDVWFGKDVWPDDEFYSHSKQKYLLTPVPNCABGCGPSWIKDGYCDKACNNSACMD	454		
QY	465	GGDCSGNSGSGRYIAGGGGTGSIIGVHPMQFGGINSVSYCNOGCANSMADKFCQACN	524		
DB	455	GGDCSGNSGSGRYIAGGGGTGSIIGVHPMQFGGINSVSYCNOGCANSMADKFCQACN	514		
QY	525	VLSGCFDAGDCQDHFHELYKYLIPNCTHYIIPKGECLPYSPFAVEAKRGVEGAYSDNP	584		
DB	515	VLSGCFDAGDCQDHFHELYKYLIPNCTHYIIPKGECLPYSPFAVEAKRGVEGAYSDNP	574		
QY	585	IIRHASIANKMTIHLIMHSGNAATTIHNLTFQONTDEEFMOQITVEYDTRGPKLNST	644		
DB	575	IIRHASIANKMTIHLIMHSGNAATTIHNLTFQONTDEEFMOQITVEYDTRGPKLNST	634		
QY	645	AQKGYENLVSPITLLPEAEILFEDIPEKGRPKFRHDVNSTRAQOEYKIPLVNISLLP	704		
DB	635	AQKGYENLVSPITLLPEAEILFEDIPEKGRPKFRHDVNSTRAQOEYKIPLVNISLLP	694		
QY	705	KNAQSLNLTDLQLEHGDITLKGYNLSKALLRSLTMSQAKIKNOAIITDETNDSLVA	764		
DB	695	KNAQSLNLTDLQLEHGDITLKGYNLSKALLRSLTMSQAKIKNOAIITDETNDSLVA	754		
QY	765	POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDOQNPLDLETTAFRVSTHTQX	824		
DB	755	POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDOQNPLDLETTAFRVSTHTQX	814		
QY	825	TIGGANTTKERPSSLIVPLESOMTKKKITGKEKENSREBNAENHIGYTEVILGKLOHY	884		
DB	815	TIGGANTTKERPSSLIVPLESOMTKKKITGKEKENSREBNAENHIGYTEVILGKLOHY	874		
QY	885	TDSYLGFLEMEKKYVPODLDEEESLKTOLAFPTDSKNTGR	925		
DB	875	TDSYLGFLEMEKKYVPODLDEEESLKTOLAFPTDSKNTGR	915		
RESULT 8					
ABW01536					
ID ABW01536 standard; protein; 1199 AA.					
XX	AC	ABW01536;			
XX	DT	15-JAN-2004 (first entry)			

XX	XX	N-acetylglucosamine-1 (GlcNAc)-phosphotransferase protein.			
XX	XX	glycoprotein; lectin; lysosomal storage disease; gastrointestinal;			
KW	KW	N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme.			
OS	XX	Unidentified.			
XX	XX	US2003124653-A1.			
PN	XX	03-JUL-2003.			
PD	XX	21-DEC-2001; 2001US-00023890.			
PF	XX	21-DEC-2001; 2001US-00023890.			
PR	XX	21-DEC-2001; 2001US-00023890.			
XX	XX	(NOVA-) NOVARTIS PHARM INC.			
PA	XX	Canfield WM;			
PI	XX	WPI; 2003-810985/76.			
XX	XX	N-PSDB; AAD62649.			
DR	XX	Claim 8; Page 10-13; 46pp; English.			
PS	XX	The present invention provides a method of producing a glycoprotein			
CC	CC	having reduced complex carbohydrates by culturing the lectin resistant			
CC	CC	mammalian cell expressing the glycoprotein. The method is useful for			
CC	CC	producing a glycoprotein with reduced complex carbohydrates for treating			
CC	CC	lysosomal storage disease. The present invention is also useful in gene			
CC	CC	therapy. The present sequence is N-acetylglucosamine (GlcNAc)-			
CC	CC	phosphotransferase protein			
XX	XX	Sequence 1199 AA;			
Query Match					
Best Local Similarity 94.4%; Score 4630; DB 7; Length 1199;					
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;					
QY	45	SRDOYHVLFDPSYRDNIAKGSFQNRCLPMPIDVYTWVNGTDLELKELOQVREQMEEQ	104		
DB	35	SRDOYHVLFDPSYRDNIAKGSFQNRCLPMPIDVYTWVNGTDLELKELOQVREQMEEQ	94		
QY	105	KAMREILGKNTTEPTKSEKQLECLTHCIKVMVLDPALPANITLKDVPSLPSFSA	164		
DB	95	KAMREILGKNTTEPTKSEKQLECLTHCIKVMVLDPALPANITLKDVPSLPSFSA	154		
QY	165	SDIFNVAKPKNSTVSVVFPDSTKDVEDAHSGLLKGNRSQVWVRGYLTDEVPGLVLM	224		
DB	155	SDIFNVAKPKNSTVSVVFPDSTKDVEDAHSGLLKGNRSQVWVRGYLTDEVPGLVLM	214		
QY	225	QDLAFLSGFPPTFKETNOUKTYLPENLSKVKLLQLYSEASVALIKLNPKDFQELNKOT	284		
DB	215	QDLAFLSGFPPTFKETNOUKTYLPENLSKVKLLQLYSEASVALIKLNPKDFQELNKOT	274		
QY	285	KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEELRYSLSIRHAPWVR	344		
DB	275	KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEELRYSLSIRHAPWVR	334		
QY	345	NIPIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIREGLSQKFIY	404		
DB	335	NIPIVTNGQIPSWMLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIREGLSQKFIY	394		
QY	405	LNDDVWFGKDVWPDDEFYSHSKQKYLLTPVPNCABGCGPSWIKGYCDKACNNSACMD	464		
DB	395	LNDDVWFGKDVWPDDEFYSHSKQKYLLTPVPNCABGCGPSWIKGYCDKACNNSACMD	454		
QY	465	GGDCSGNSGSGRYIAGGGGTGSIIGVGHPMQFGGINSVSYCNOGCANSMADKFCQACN	524		

Db 455 GEDCSGNSGSRVYIAGGGTGSIGVGQPMQFGGINSVSYCNQGCANSLADKPCDQACN 514

QY 525 VLSGCGDADCCGDHHEHLYKYLILPNQHYIIPKGECLPYSPFAVAKRGVAGYSPNP 584

Db 515 VLSGCGDADCCGDHHEHLYKYLILPNQHYIIPKGECLPYSPFAVAKRGVAGYSPNP 574

QY 585 IIRHASIAKWKTIHILHMSGNATIIHNLTFQNTNDEFFMQITVEVDTEGPKLNT 644

Db 575 IIRHASIAKWKTIHILHMSGNATIIHNLTFQNTNDEFFMQITVEVDTEGPKLNT 634

QY 645 AOKGVNLSPTILLPEALILFEDIPKRPKPKRHDVNSRRRAQEEVKIPLVNIISLP 704

Db 635 AOKGVNLSPTILLPEALILFEDIPKRPKPKRHDVNSRRRAQEEVKIPLVNIISLP 694

QY 705 KQALSLNTLDLQLEHGDITLKQYNSKSLRSPFMSQAKIKNOAIITDETDSLVA 764

Db 695 KQALSLNTLDLQLEHGDITLKQYNSKSLRSPFMSQAKIKNOAIITDETDSLVA 754

QY 765 POKQVHKSILPNSLGVSRQLQTLFPVAVSVKNGHDOGNPLDLETTARFVETHTQK 824

Db 755 POKQVHKSILPNSLGVSRQLQTLFPVAVSVKNGHDOGNPLDLETTARFVETHTQK 814

QY 825 TIGGANTKPKPSLIVPESOMTKKTIKGEKENSREBNENHIGTEVILGKQOHY 884

Db 815 TIGGANTKPKPSLIVPESOMTKKTIKGEKENSREBNENHIGTEVILGKQOHY 874

QY 885 TDSYLGFLPWEKKYFODLLDEESLKTQLAYFTDSKNTGR 925

Db 875 TDSYLGFLPWEKKYFODLLDEESLKTQLAYFTDSKNTGR 915

RESULT 9

AAE25294 standard; protein; 1196 AA.

XX AAE25294;

AC 30-OCT-2002 (first entry)

XX

DE Human nucleic acid-associated protein (NAAP-13).

XX

XX Human; nucleic acid-associated protein; NAAP-13; neurological disorder; arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis; lymphoma; epilepsy; Alzheimer's disease; developmental; anticomulant; autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania; gene therapy; neotropic; neuroprotective; cerebroprotective; virucide; immunosuppressive; protozoacide; antimicrobial.

XX

OS Homo sapiens.

XX

PH Key location/Qualifiers

FT Peptide 1..34

FT Domain /label= Signal peptide

FT Domain 1..19

FT Domain /note= "Cytosolic domain"

FT Domain 20..42

FT Protein /note= "Transmembrane domain"

FT Domain 35..1196

FT Domain /note= "Mature human NAAP-13"

FT Domain 43..1152

FT Domain /note= "Non-cytosolic domain"

FT Domain 1153..1175

FT Domain /note= "Transmembrane domain"

FT Domain 1176..1196

FT Domain /note= "Cytosolic domain"

XX

PN WO200250279-A2.

XX

PD 27-JUN-2002.

XX

PF 19-DEC-2001; 2001WO-US050256.

XX

PR 21-DEC-2000; 2000US-0257714P.

PR 05-JAN-2001; 2001US-0260081P.

PR 16-JAN-2001; 2001US-0262302P.

PR 23-JAN-2001; 2001US-0263823P.

PR 02-FEB-2001; 2001US-0266088P.

PR 29-OCT-2001; 2001US-0348442P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Baughin MR, Lu Y, Arvizu C, Rankumar J, Yao MG, Polickly JL;

PI Wala NK, Tribouley KM, Yue H, Batta S, Ding L, Lal PG;

PI Borowicz MU, Lu DM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;

PI Gietzen KJ, Tang YT, Warren BA, Mason PM, Barford N, Hatfield AA;

PI Lee SA, Yang J, Gorvad AE, Emerling BM, Marguis JP, Lee SY;

PI Swarnakar A, Reddy R;

XX

DR WPI: 2002-519887/55.

DR N-PSDB: AAD41203.

XX

PT Nucleic acid associated proteins and nucleic acids for diagnosing,

PT treating and preventing cell proliferative (e.g. cancers), neurological

PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).

XX

PS Claim 68; Page 168-172; 193pp; English.

XX

CC The invention relates to nucleic acid-associated proteins (NAAP) and

CC nucleic acids. The nucleic acid and amino acid sequences are useful for

CC diagnosing, treating and preventing cell proliferative e.g.

CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological

CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and

CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections

CC (e.g. malaria, or leishmania), as well as in assessing the effects of

CC exogenous compound on the expression of nucleic acid and amino acid

CC sequences of nucleic acid-associated proteins. The invention is useful in

CC gene therapy. The present sequence is human NAAP-13

XX

SQ Sequence 1196 AA;

QY 1 MFRLKORQYTLSTRYGLVYCFGLVVTIVSAFOGFEVLEMSRDQYHVFDSYRNI 60

Db 1 MFRLKORQYTLSTRYGLVYCFGLVVTIVSAFOGFEVLEMSRDQYHVFDSYRNI 60

QY 61 AGKSPFNRLCLPMPIDVVTWNGTDLLEKELQVREQMESEOKAMEIILKNTTEPTK 120

Db 61 AGKSPFNRLCLPMPIDVVTWNGTDLLEKELQVREQMESEOKAMEIILKNTTEPTK 120

QY 121 KSEKOLECLLTHGICKPMVLDPALPANTTLKDVSLVSPHSASDIFVAKPKPSTNV 180

Db 121 KSEKOLECLLTHGICKPMVLDPALPANTTLKDVSLVSPHSASDIFVAKPKPSTNV 180

QY 181 SVVVPDSTVDVEDAHSGLLKGNRSQTWRGYLTTRKEVGLVLMODLAFLSGPPPTKET 240

Db 181 SVVVPDSTVDVEDAHSGLLKGNRSQTWRGYLTTRKEVGLVLMODLAFLSGPPPTKET 240

QY 241 NQKTLPEMLSSKVKYLQVSEASVALKLNPPDPOELNKOTKNTTICKETLSPA 300

Db 241 NQKTLPEMLSSKVKYLQVSEASVALKLNPPDPOELNKOTKNTTICKETLSPA 300

QY 301 YLLMDLSAISQSKODEDISASRFEDNEBELRSLRSIBHAPVWNIPIVTNQIIPSWNL 360

Db 301 YLLMDLSAISQSKODEDISASRFEDNEBELRSLRSIBHAPVWNIPIVTNQIIPSWNL 360

QY 361 DNEPRTIVHGVFRNLHLPTFSSGPAIESHTRIEGSLQKTIYANDVMFGKWPPDF 420

Db 361 DNEPRTIVHGVFRNLHLPTFSSGPAIESHTRIEGSLQKTIYANDVMFGKWPPDF 420

QY 421 YSHSKQXYVLTWPVNCAGCGPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYLAG 480

Db 421 YSHSKQXYVLTWPVNCAGCGPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYLAG 480

434

QY 481 GGGTSGIGVGHFMWFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCFDAGDCGDHF 540
DB 435 -----KGGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGDCGDHF 480
QY 541 HELYKVILLPNQTHYIIPKGECLPYSPFAEVAKRGVAGVSDNPIIRHASTANKKTIHL 600
DB 481 HELYKVILLPNQTHYIIPKGECLPYSPFAEVAKRGVAGVSDNPIIRHASTANKKTIHL 540
QY 601 IMHSGMATTIHNLTFQNTNDEEFMQITVEVDTRBEGKLNSTAKGYENLVSPITLLP 660
DB 541 IMHSGMATTIHNLTFQNTNDEEFMQITVEVDTRBEGKLNSTAKGYENLVSPITLLP 600
QY 661 EAILFEDIPKEKRPFKRHDVNSTRAQOEVKIPLVNISLLPDAQSLNTLDLQLEH 720
DB 601 EAILFEDIPKEKRPFKRHDVNSTRAQOEVKIPLVNISLLPDAQSLNTLDLQLEH 660
QY 721 GDITLKYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLAPQEKVHKSILPNSLG 780
DB 661 GDITLKYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLAPQEKVHKSILPNSLG 720
QY 781 VSERLQRLTFPAVSVKVNCHDQGNPDLLETTAFRVETHQKTIGAVTKEKPPSLIV 840
DB 721 VSERLQRLTFPAVSVKVNCHDQGNPDLLETTAFRVETHQKTIGAVTKEKPPSLIV 780
QY 841 PLESQWTEKKITTGKEKENSMEENAHNIGVTEVLLGRKLOHYTDSYLGFLEMEKKYF 900
DB 781 PLESQWTEKKITTGKEKENSMEENAHNIGVTEVLLGRKLOHYTDSYLGFLEMEKKYF 840
QY 901 QDLDBESLKTQLAYFTDSCNTGRQK 928
DB 841 QDLDBESLKTQLAYFTDSCNTGRQK 900
RESULT 10
ADA54824
ID ADA54824 standard, protein, 846 AA.
AC ADA54824;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 2392.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nocitropic;
KW Gene therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
OS Homo sapiens.
XX
FN EPI293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
XX
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
PI Isogai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S,
PI Yamamoto U, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
DR WPI; 2003-395539/38.
XX
DR N-PSDB; ADA53185.
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 14; SEQ ID NO 2392; 205bp; English.

XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 846 AA;
Query Match 91.1%; Score 4468; DB 6; Length 846;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 844; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLFRLQRLTFPAVSVKVNCHDQGNPDLLETTAFRVETHQKTIGAVTKEKPPSLIV 60
DB 1 MLFRLQRLTFPAVSVKVNCHDQGNPDLLETTAFRVETHQKTIGAVTKEKPPSLIV 60
QY 61 AGKSFQNRCLPMPIDVYVWNGTDLLEIKELQVREQMEBEOKAMREILGNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVWNGTDLLEIKELQVREQMEBEOKAMREILGNTTEPTK 120
QY 121 KSEKQECCLTTCICIKPMLVLPALPANITLKQVBSLYSFHSASDIENVAKPXNPSTV 180
DB 121 KSEKQECCLTTCICIKPMLVLPALPANITLKQVBSLYSFHSASDIENVAKPXNPSTV 180
QY 181 SVVVFSDTKQVEDAHSGILKGNRQTVWRGYLTTFDKXVPGLVIMODLAFSGPPEFKEF 240
DB 181 SVVVFSDTKQVEDAHSGILKGNRQTVWRGYLTTFDKXVPGLVIMODLAFSGPPEFKEF 240
QY 241 NQKTKLPENLSKRYKLLQYSPASVALLKLNPKDFQELNKKYKXNTIDGKELTISPA 300
DB 241 NQKTKLPENLSKRYKLLQYSPASVALLKLNPKDFQELNKKYKXNTIDGKELTISPA 300
QY 301 YLLMDLSAISQKODEISARFEDNBEALYSLSIERHAPWRNFIYVNGQIPSMNL 360
DB 301 YLLMDLSAISQKODEISARFEDNBEALYSLSIERHAPWRNFIYVNGQIPSMNL 360
QY 361 DNPRVTVTHQDVFRNLSHLPFTSSPAISHIRIGLSQKFIYLLNDVWFGDVPDDF 420
DB 361 DNPRVTVTHQDVFRNLSHLPFTSSPAISHIRIGLSQKFIYLLNDVWFGDVPDDF 420
QY 421 YSHSKGQKYLITWVPNCABGCGSWIKDGYCKKACNNSACMDWDGDCGNSGGSRYIAG 480
DB 421 YSHSKGQKYLITWVPNCABGCGSWIKDGYCKKACNNSACMDWDGDCGNSGGSRYIAG 480
QY 481 GGGTSGIGVGHFMWFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCFDAGDCGDHF 540
DB 481 GGGTSGIGVGHFMWFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCFDAGDCGDHF 540
QY 541 HELYKVILLPNQTHYIIPKGECLPYSPFAEVAKRGVAGVSDNPIIRHASTANKKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYSPFAEVAKRGVAGVSDNPIIRHASTANKKTIHL 600
QY 601 IMHSGMATTIHNLTFQNTNDEEFMQITVEVDTRBEGKLNSTAKGYENLVSPITLLP 660
DB 601 IMHSGMATTIHNLTFQNTNDEEFMQITVEVDTRBEGKLNSTAKGYENLVSPITLLP 660
QY 661 EAILFEDIPKEKRPFKRHDVNSTRAQOEVKIPLVNISLLPDAQSLNTLDLQLEH 720
DB 661 EAILFEDIPKEKRPFKRHDVNSTRAQOEVKIPLVNISLLPDAQSLNTLDLQLEH 720
QY 721 GDITLKYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLAPQEKVHKSILPNSLG 780
DB 721 GDITLKYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLAPQEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVNCHDQGNPDLLETTAFRVETHQKTIGAVTKEKPPSLIV 840
DB 781 VSERLQRLTFPAVSVKVNCHDQGNPDLLETTAFRVETHQKTIGAVTKEKPPSLIV 840
QY 841 PLESQW 846
DB 841 PLESQW 846

RESULT 11
 ID ABU07381 standard; protein; 1459 AA.
 XX
 AC ABU07381;
 XX
 DT 28-JAN-2003 (first entry)
 XX
 DE Human protein NOV12.
 XX
 KW Human; NOVX; cardiomyopathy; atherosclerosis; cancer; hypertension;
 KW diabetes; inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
 KW immunoglobulin A nephropathy; cirrhosis; arthritis; Alzheimer's disease;
 KW Parkinson's disease; goitre; infection; stroke; muscular dystrophy;
 KW epilepsy; wasting disorder; neurogenesis; cell differentiation;
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis;
 KW gene therapy; single nucleotide polymorphism; SNP.
 XX
 OS Homo sapiens.
 XX
 PN WO200285922-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 11-APR-2002; 2002WO-US011634.
 XX
 PR 23-APR-2001; 2001US-0285748P.
 PR 24-APR-2001; 2001US-0286068P.
 PR 25-APR-2001; 2001US-0286292P.
 PR 03-MAY-2001; 2001US-0286334P.
 PR 16-MAY-2001; 2001US-0291241P.
 PR 14-SEP-2001; 2001US-0322284P.
 XX
 PA (CURA-) CURAGEN CORP.
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Pena CE, Guo X, Shinkets RA, Padigaru M, Kekuda R, Spytek KA,
 PI Mehreban F, Topper JN, Malayanar UM, Wasserman S, Edinger S,
 PI Smithson G, Gunther E, Komuves L;
 XX
 DR WPI; 2003-058712/05.
 DR N-PSDB; ABX10235.
 XX
 PT New NOVX polypeptides and nucleic acids, useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 PT
 XX
 PS Claim 1; Page 94-95; 301pp; English.
 CC
 CC The invention relates to an isolated polypeptide comprising any one of 17
 CC human NOVX (1-9, 10a, 10b, 11-16) appearing as ABU07389-ABU07385, a
 CC mature form of 1c, or a variant of them, where one or more residues of
 CC the variant differs in not more than 15 % from the residues of the
 CC sequence of them and their encoding polynucleotides appearing as ABX10223
 CC -ABX10239. Also included are NOVX expression vectors, transformed cells,
 CC antibodies, identifying an agent that binds to or modulates the
 CC expression or activity of NOVX and screening for a modulator of activity
 CC or of latency or predisposition to a NOVX-associated disorder. The NOVX
 CC polypeptides, polynucleotides and antibodies are useful in manufacturing
 CC a medicament for treating or preventing a syndrome associated with NOVX-
 CC associated disorder, such as cardiomyopathy, atherosclerosis, cancer,
 CC hypertension, diabetes, inflammation, autoimmune disorders, allergies,
 CC blood disorders, acquired immunodeficiency syndrome (AIDS), obesity,
 CC asthma, immunoglobulin (Ig) A nephropathy, cirrhosis, arthritis,
 CC Alzheimer's disease, Parkinson's disease, goitre, infections (e.g.
 CC bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy, and
 CC other wasting disorders associated with chronic diseases. The nucleic
 CC acids and polypeptides may also be used as targets for the identification
 CC of small molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and

CC angiogenesis, in gene therapy, in generation of antibodies that bind
 CC immunospecifically to NOVX substances for use in therapeutic or
 CC diagnostic methods. The nucleic acids are further used as hybridisation
 CC probes, in chromosome mapping, tissue typing, preventive medicine, and
 CC pharmacogenomics. The polypeptides are also useful as vaccines. The
 CC present sequence represents a NOVX protein
 CC
 XX
 SO Sequence 1459 AA;
 Query Match 89.5%; Score 4391; DB 6; Length 1459;
 Best Local Similarity 73.0%; Pred. No. 0;
 Matches 884; Conservative 13; Mismatches 28; Indels 286; Gaps 13;
 QY 1 MFKKLQRTYTCLSRRYGLVYGLVNTTVISAFQGE----- 39
 Db 1 MFKKLQRTYTCLSRRYGLVYGLVNTTVISAFQGEVBARPAKPIVHRAPTTK 60
 QY 40 ----- 39
 Db 61 NHPAQNVDAAVEKSGIRRGKGRRAVSLQDMPGTRGCANFPACFDCKEVSQKRL 120
 QY 40 -----VLEMSRDQYVHLPDSYED 58
 Db 121 YILQNGHMLTDGWTWLLHSGLLIQGPASRGCVLLKAVVLEMSRDQYVHLPDSYED 180
 QY 59 NIKGSFQRLCLPPIIDVYTVWNGTDLELKELOQVREOMEEOKANRETLGKTTPE 118
 Db 181 NIKGSFQRLCLPPIIDVYTVWNGTDLELKELOQVREOMEEOKANRETLGKTTPE 240
 QY 119 TKSEKQECLELHCICKVEMLVDPALPANITLKDVPSPISPEFHSASDIFNAKRNPT 178
 Db 241 TKSEKQECLELHCICKVEMLVDPALPANITLKDVPSPISPEFHSASDIFNAKRNPT 300
 QY 179 NVSVVVPDSTKD-----VEDHSGLLKGN-----ROTW----- 208
 Db 301 NVSVVVPDSTKDGLTQKVTPEWKCCEGEVSNANIMWKTDLGSPRRPLPMPVLEPR 360
 QY 209 -----RGYLTDEKVEPGLV- MODLAFSGPP 234
 Db 361 AQSSALQITPRVSQDPAANSYELKDTPLIRGTA---KEVPBETGLQPLSFAHCLP 417
 QY 235 -----PTF---KET-----NOLKT---KL 247
 Db 418 TDLARKNELRDEVKMYKODPSILHTKETCFPLREQVESGSEYKSEENIKELTKGSKV 477
 QY 248 PEN-----LSKKYKQLQYS-----EASVALKLNNPKDFOELNKQTK 285
 Db 478 EENITSDLESSESDLEIDNEAVIEPTDPSPEMGDGEASVALKLNNPKDFOELNKQTK 537
 QY 286 KMTTIDKEELTISPAVILMDLSAISQSKODEDISARFEDNEELRYSLSIRRAPWYN 345
 Db 538 KMTTIDKEELTISPAVILMDLSAISQSKODEDISARFEDNEELRYSLSIRRAPWYN 597
 QY 346 IFVTVNGQIPSWLNDNPRVTVTHQDVRANLSHLPTESSPAIBSHIRIBGLSOKFTYL 405
 Db 598 IFVTVNGQIPSWLNDNPRVTVTHQDVRANLSHLPTESSPAIBSHIRIBGLSOKFTYL 657
 QY 406 NDVVMGKXVWPDPDFSHSKGQKVYLTWVPNCAEGCPGSMWKDGYCDKACNNSACMDG 465
 Db 658 NDVVMGKXVWPDPDFSHSKGQKVYLTWVPNCAEGCPGSMWKDGYCDKACNNSACMDG 717
 QY 466 GDCSGNSGSGSRITAGGGGTGSGVGHPMOFGGINSVSYCNQGCANSLADFCQACNV 525
 Db 718 GDCSGNSGSGSRITAGGGGTGSGVGHPMOFGGINSVSYCNQGCANSLADFCQACNV 777
 QY 526 LSCGFPAGDQO-----DHFEHYKYLILPNQTHYII 557
 Db 778 LSCGFPAGDQO-----DHFEHYKYLILPNQTHYII 837
 QY 558 PKGECIPYSPFAVARGVEGAYSDNPITIRASINNKTKTILIMHSGNNAATTIHPNLT 617
 Db 838 PKGECIPYSPFAVARGVEGAYSDNPITIRASINNKTKTILIMHSGNNAATTIHPNLT 897


```

RESULT 13
ADD27817
ID ADD27817 standard; protein; 908 AA.
XX
AC ADD27817;
XX
DT 15-JAN-2004 (first entry)
XX
DE GlcNAc-phosphotransferase associated protein #1.
XX
KM mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase;
KM UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.
XX
OS Mus musculus.
XX
PN US2003119088-A1.
XX
PD 26-JUN-2003.
XX
PF 21-DEC-2001; 2001US-00023888.
XX
PR 21-DEC-2001; 2001US-00023888.
XX
PA (NOVA-) NOVAAZYME PHARM INC.
XX
PI Canfield W, Kudo M;
XX
DR MPI; 2003-801323/75.
XX
PT Phosphorylating a protein for treating a patient suffering from a
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
PT protein.
XX
PS Disclosure; SEQ ID NO 9; 55pp; English.
XX
CC The invention relates to a method of phosphorylating a protein comprising
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of a GlcNAc-phosphotransferase associated protein.
XX
SQ Sequence 908 AA;

Query Match 79.4%; Score 3894; DB 7; Length 908;
Best Local Similarity 79.9%; Pred. No. 3,6e-305;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MFLKLRQRTYTCLSHRVGLVYCFVGAVVTVISAFQFGEVWLEMSRDQVHLPFSYRNI 60
DB 1 MFLKLRQRTYTCLSHRVGLVYCFVGAVVTVISAFQFGEVWLEMSRDQVHLPFSYRNI 60
QY 61 AKSFFQNRICLPMPIDIVVTWNGTDLLELKELOQVRBQMEBQKAMEIIGKNTTEPTK 120
DB 61 AKSFFQNRICLPMPIDIVVTWNGTDLLELKELOQVRBQMEBQKAMEIIGKNTTEPTK 120
QY 61 AKSFFQNRICLPMPIDIVVTWNGTDLLELKELOQVRBQMEBQKAMEIIGKNTTEPTK 120
DB 61 AKSFFQNRICLPMPIDIVVTWNGTDLLELKELOQVRBQMEBQKAMEIIGKNTTEPTK 120
QY 121 KSEKOLECLTHCIVPMVMDPALPANITLKDVPASDIPSHASDINVAKKPNPSTNV 180
DB 121 KSEKOLECLTHCIVPMVMDPALPANITLKDVPASDIPSHASDINVAKKPNPSTNV 180
QY 121 KSEKOLECLTHCIVPMVMDPALPANITLKDVPASDIPSHASDINVAKKPNPSTNV 180
DB 121 KSEKOLECLTHCIVPMVMDPALPANITLKDVPASDIPSHASDINVAKKPNPSTNV 180
QY 181 SYVVPDSTKDVADHAGLLKGNRSQRTWRGVLTDDKEVPGVLMQDLAFISGPPPTKET 240
DB 181 SYVVPDSTKDVADHAGLLKGNRSQRTWRGVLTDDKEVPGVLMQDLAFISGPPPTKET 240
QY 181 SYVVPDSTKDVADHAGLLKGNRSQRTWRGVLTDDKEVPGVLMQDLAFISGPPPTKET 240
DB 181 SYVVPDSTKDVADHAGLLKGNRSQRTWRGVLTDDKEVPGVLMQDLAFISGPPPTKET 240
QY 241 NOLTKRLPEN-LSSKVKLLQLYSEASVALLKLNKXQDQELNKQTKKMTIDGELTISP 299
DB 241 NOLTKRLPEN-LSSKVKLLQLYSEASVALLKLNKXQDQELNKQTKKMTIDGELTISP 299
QY 241 NOLTKRLPEN-LSSKVKLLQLYSEASVALLKLNKXQDQELNKQTKKMTIDGELTISP 299
DB 241 NOLTKRLPEN-LSSKVKLLQLYSEASVALLKLNKXQDQELNKQTKKMTIDGELTISP 299
QY 300 AVILMDLSAISQKODEDISASRFEDNEBELRYSLSIERHAPWYANIFIVTNGQIPSWLN 359
DB 300 AVILMDLSAISQKODEDISASRFEDNEBELRYSLSIERHAPWYANIFIVTNGQIPSWLN 359
QY 301 ATLMWDLAISQKODEDASASRFEDNEBELRYSLSIERHAPWYANIFIVTNGQIPSWLN 360
DB 301 ATLMWDLAISQKODEDASASRFEDNEBELRYSLSIERHAPWYANIFIVTNGQIPSWLN 360

```

```

QY 360 LDNPRVITVHODVFENLSLHPFSSPAIESHHRFEGSOKFIYLANDVDFGKDVPPD 419
DB 361 LDNPRVITVHODVFENLSLHPFSSPAIESHHRFEGSOKFIYLANDVDFGKDVPPD 420
QY 420 FYSHSKGQKYLLTWVPNCAEGCGPGSMIKDGYCDKACNNSACDWDGDCSGSGGRYTA 479
DB 421 FYSHSKGQKYLLTWVPNCAEGCGPGSMIKDGYCDKACNNSACDWDGDCSGSGGRYTA 480
QY 480 GGGGTSGIGVHWMQGGGINSVYNGGACNSMTLADKFCDDACNVLSGFGPAGCGGDH 539
DB 481 RGGGTGNTGAGHWMQGGGINSVYNGGACNSMTLADKFCDDACNVLSGFGPAGCGGDH 540
QY 540 FHELYKVILLPNQTHYIIPKBCLPYFSFAEVAKGVGEVAGSDNPILIRASIANKWKTH 599
DB 541 FHELYKVILLPNQTHYIIPKBCLPYFSFAEVAKGVGEVAGSDNPILIRASIANKWKTH 600
QY 600 LHHSGMNAITTHFNITPQNTDDEPRMOITVEVDTREPKLNSTQKGYENLVSPITLL 659
DB 601 LHHSGMNAITTHYFNITLQNDDEPRKIQIAVEVDTREPKLNSTQKAYESLVSPVPL 660
QY 660 PEALIFEDIPEKREPEKFRADVNSTRAQAEVKKIPLVNISSLPRDAQSLNTLDLOLE 719
DB 661 PQAADVPFEDVPKREKPRKIRRDVATGSGFQEEVKIPRNISLIPKRAQVRLSNDILOE 720
QY 720 HGDITLKGNLSKALLRSFLMNSQHAKIKNOAITDETNSLVAPQEKQVHKSILPNSL 779
DB 721 RGDITLKGNLSKALLRSFLMNSQHAKIKNOAITDETNSLVAPQEKQVHKSILPNSL 777
QY 780 GVSERLQRLPPVAVSKVNGHOGQNPPLDLETTAFRVEHTQKTIIGANVTEKPPSLI 839
DB 778 AGHREBRTAPAEITYVAGRDHALNPVLETNARL-----AQPLTGVTVKEMLSPLI 832
QY 840 VPLESQTKKEKKITGKEKENSMEENAEHNVGTEVLLGRKLOHYTDSYGLFPMKKKY 899
DB 843 VPESHLP-----KEESDRAEGNA--VPKEHVPGRRLQ--QNVPGFLPWEKKKY 879
QY 900 FQDLDEESLKTQLAFTDSKNTGROLK 928
DB 880 FQDLDEESLKTQLAFTDRKHTGRQLK 908

RESULT 14
ABM01491
ID ABM01491 standard; protein; 908 AA.
XX
AC ABM01491;
XX
DT 15-JAN-2004 (first entry)
XX
DE Mouse protein #1 used to illustrate the method of the invention.
XX
KM Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KM lysosomal storage disease; gastrointestinal; mouse;
KM lectin resistant cell; deoxymannojirimycin; kifunensine;
KM glycosylation inhibition.
XX
OS Mus musculus.
XX
PN US2003124652-A1.
XX
PD 03-JUL-2003.
XX
PF 21-DEC-2001; 2001US-00023889.
XX
PR 21-DEC-2001; 2001US-00023889.
XX
PA (NOVA-) NOVAAZYME PHARM INC.
XX
PI Canfield WM;
XX
DR MPI; 2003-810984/76.
XX

```


PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
XX
XX Disclosure; Page 23-25; 46pp; English.
PS
XX The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is mouse protein used to
CC illustrate the method of the invention
XX
XX Sequence 908 AA;

Query Match 79.4%; Score 3894; DB 7; Length 908;

Best Local Similarity 79.9%; Pred. No. 3.6e-305; Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MFLKLRQRTYCLSHRYGLVYCFGLGVVTVTISAFQFGEVVLWMSRDQYHLPDSYRDN 60
DB 1 MFLKLRQRTYCLSHRYGLVYCFGVVTVTISAFQFGEVVLWMSRDQYHLPDSYRDN 60
QY 61 AGKSFQNRCLPMPIDVYVYTWNGTDLLELKELOQVREHMEBQKAMRETLGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVYTWNGTDLLELKELOQVREHMEBQKAMRETLGKNTTEPTK 120
QY 121 KSEKOLECHLTHCIVPMVLVDPALPANITLKDVPSLYPSHSADIFNVAKPKNPSTNV 180
DB 121 KSEKOLECHLTHCIVPMVLVDPALPANITLKDVPSLYPSHSADIFNVAKPKNPSTNV 180
QY 122 KSEKOLECHLTHCIVPMVLVDPALPANITLKDVPSLYPSHSADIFNVAKPKNPSTNV 180
DB 122 KSEKOLECHLTHCIVPMVLVDPALPANITLKDVPSLYPSHSADIFNVAKPKNPSTNV 180
QY 181 SVVYVDSSTVDVDAHSGLLKNSRQTVWRGVTITTKKAVPGVLMODLAFLSGFPPTKET 240
DB 181 SVVYVDSSTVDVDAHSGLLKNSRQTVWRGVTITTKKAVPGVLMODLAFLSGFPPTKET 240
QY 181 PVVYVDSSTVDVDAHSGLLKNSRQTVWRGVTITTKKAVPGVLMODLAFLSGFPPTKET 240
DB 181 PVVYVDSSTVDVDAHSGLLKNSRQTVWRGVTITTKKAVPGVLMODLAFLSGFPPTKET 240
QY 241 NOLKTKLPEN-LSKVKTLQLYSEASVALKLKLNPKDFOELNKQTKKMTIDGKELTISP 299
DB 241 NOLKTKLPEN-LSKVKTLQLYSEASVALKLKLNPKDFOELNKQTKKMTIDGKELTISP 299
QY 241 SOLKTKLPKAPLAKIKLRLYSEASVALKLKLNPKDFOELNKQTKKMTIDGKELTISP 300
DB 241 SOLKTKLPKAPLAKIKLRLYSEASVALKLKLNPKDFOELNKQTKKMTIDGKELTISP 300
QY 300 AYLLWDLAISQSKODEDISASRPFNDNEELRYSLRSIRHAPWVNFIITNGQIPSWLN 359
DB 300 AYLLWDLAISQSKODEDISASRPFNDNEELRYSLRSIRHAPWVNFIITNGQIPSWLN 359
QY 301 AYLLWDLAISQSKODEDISASRPFNDNEELRYSLRSIRHAPWVNFIITNGQIPSWLN 360
DB 301 AYLLWDLAISQSKODEDISASRPFNDNEELRYSLRSIRHAPWVNFIITNGQIPSWLN 360
QY 360 LNPRTVITVHODVFRNLSHLPTSPSPALSHIHEGLSKOFTYLVNDVWFGKDVWPD 419
DB 360 LNPRTVITVHODVFRNLSHLPTSPSPALSHIHEGLSKOFTYLVNDVWFGKDVWPD 419
QY 361 LNPRTVITVHODVFRNLSHLPTSPSPALSHIHEGLSKOFTYLVNDVWFGKDVWPD 420
DB 361 LNPRTVITVHODVFRNLSHLPTSPSPALSHIHEGLSKOFTYLVNDVWFGKDVWPD 420
QY 420 FYSHSKGQKYLITWVPNCABGCPGSMIKDGYCDKACNNSACDMWDGDCSGNSGGSRYIA 479
DB 420 FYSHSKGQKYLITWVPNCABGCPGSMIKDGYCDKACNNSACDMWDGDCSGNSGGSRYIA 479
QY 421 FYSHSKGQKYLITWVPNCABGCPGSMIKDGYCDKACNNSACDMWDGDCSGNSGGSRYIA 480
DB 421 FYSHSKGQKYLITWVPNCABGCPGSMIKDGYCDKACNNSACDMWDGDCSGNSGGSRYIA 480
QY 480 GGGGSGSIVGHPMGQGGINSVSYCNGGCANSLWADFCDOACNVLSGCPDAGCCGDH 539
DB 480 GGGGSGSIVGHPMGQGGINSVSYCNGGCANSLWADFCDOACNVLSGCPDAGCCGDH 539
QY 481 RGGGSGSIVGHPMGQGGINSVSYCNGGCANSLWADFCDOACNVLSGCPDAGCCGDH 540
DB 481 RGGGSGSIVGHPMGQGGINSVSYCNGGCANSLWADFCDOACNVLSGCPDAGCCGDH 540
QY 540 FHELYVILLPQNTYIIPKGECLPFSPFAVAKRGEVGAASDNIIRHSAIANMKTIH 599
DB 540 FHELYVILLPQNTYIIPKGECLPFSPFAVAKRGEVGAASDNIIRHSAIANMKTIH 599
QY 541 FHELYVILLPQNTYIIPKGECLPFSPFAVAKRGEVGAASDNIIRHSAIANMKTIH 600
DB 541 FHELYVILLPQNTYIIPKGECLPFSPFAVAKRGEVGAASDNIIRHSAIANMKTIH 600
QY 600 LIMPGENATITTYFNLTIONANDEBKQIAVEVDTREAPKLNSTTORAAYESLVSPVPL 660
DB 600 LIMPGENATITTYFNLTIONANDEBKQIAVEVDTREAPKLNSTTORAAYESLVSPVPL 660
QY 660 PEALIFEDIPKRRPKFKRDVNSTRAQSEVNIPLVNIISLPKQOLSLNTDLE 719
DB 660 PEALIFEDIPKRRPKFKRDVNSTRAQSEVNIPLVNIISLPKQOLSLNTDLE 719
QY 661 PQADVFEVPEKERRPKIRHVDVATGFOBEVNIIPRNIISLPKQOLSLNTDLE 720
DB 661 PQADVFEVPEKERRPKIRHVDVATGFOBEVNIIPRNIISLPKQOLSLNTDLE 720
QY 720 HGDITLKGYLSKALRSFLMNSQHAKIKNOAITTDETNSLVAPQEKVHKSLIPNSL 779
DB 720 HGDITLKGYLSKALRSFLMNSQHAKIKNOAITTDETNSLVAPQEKVHKSLIPNSL 779

DB 721 RGDITLKGYLSKALRSFLMNSQHAKIKNOAITTDETNSLVAPQEKVHKSLIPNSL 779
QY 780 GVSERLQRTYCLSHRYGLVYCFGLGVVTVTISAFQFGEVVLWMSRDQYHLPDSYRDN 839
DB 778 AGHRSERWTAPAEITVTKGRDHALNPPEVLETNARL-----AQPTLGVTSEKENSPLI 832
QY 840 VPLESQMTKEKKTIGKEKENSRYEENAHIGTETVLGRKLOHYTDSYLGFLPWEKKY 899
DB 833 VPESHLP-----KEESDRAEGNA---VFKSLVPRRLD---QNTPLPWEKKY 879
QY 900 FQDLDEESLTKQLAAYFTDSKNTGSLK 928
DB 880 FQDLDEESLTKQLAAYFTDSKNTGSLK 908

RESULT 15

ABW01540 ID ABW01540 standard; protein; 908 AA.

XX AC ABW01540;

XX DT 15-JAN-2004 (first entry)

XX DE Mouse protein #1 used to illustrate the method of the invention.

XX KM Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;

XX KM N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse.

XX OS Mus musculus.

XX FN US2003124653-A1.

XX PD 03-JUL-2003.

XX PF 21-DEC-2001; 2001US-00023890.

XX PR 21-DEC-2001; 2001US-00023890.

XX PA (NOVA-) NOVAZYME PHARM INC.

XX PI Canfield WM;

XX DR WPI; 2003-810985/76.

XX PT Producing a glycoprotein with reduced complex carbohydrates by culturing
PT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.

XX PS Disclosure; Page 23-25; 46pp; English.

XX CC The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing the lectin resistant
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is mouse protein used to illustrate the
CC method of the invention
XX
XX Sequence 908 AA;

QY Query Match 79.4%; Score 3894; DB 7; Length 908;

Best Local Similarity 79.9%; Pred. No. 3.6e-305; Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MFLKLRQRTYCLSHRYGLVYCFGLGVVTVTISAFQFGEVVLWMSRDQYHLPDSYRDN 60
DB 1 MFLKLRQRTYCLSHRYGLVYCFGVVTVTISAFQFGEVVLWMSRDQYHLPDSYRDN 60
QY 61 AGKSFQNRCLPMPIDVYVYTWNGTDLLELKELOQVREHMEBQKAMRETLGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVYTWNGTDLLELKELOQVREHMEBQKAMRETLGKNTTEPTK 120

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:06:34 ; Search time 26.1576 Seconds

(without alignments)
1831.549 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907
Sequence: 1 MLFLLQRQTYTCLSHRYGL.....SLKTQLAYFTDSKNTGRQLK 928Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pdp:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pdp:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pdp:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pdp:*
5: /cgn2_6/ptodata/2/1aa/PTCUTS.COMB.pdp:*
6: /cgn2_6/ptodata/2/1aa/backfilest.pdp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4907	100.0	928	4	US-09-635-872A-1 Sequence 1, Appli
2	4907	100.0	928	4	US-09-636-077A-1 Sequence 1, Appli
3	4907	100.0	928	4	US-09-636-060C-1 Sequence 1, Appli
4	4907	100.0	928	4	US-09-986-552-1 Sequence 1, Appli
5	3894	79.4	908	4	US-09-635-872A-15 Sequence 15, Appli
6	3894	79.4	908	4	US-09-636-077A-15 Sequence 15, Appli
7	3894	79.4	908	4	US-09-636-060C-15 Sequence 15, Appli
8	3894	79.4	908	4	US-09-986-552-15 Sequence 15, Appli
9	503	10.3	113	4	US-09-635-872A-11 Sequence 11, Appli
10	503	10.3	113	4	US-09-636-060C-11 Sequence 11, Appli
11	503	10.3	113	4	US-09-636-060C-11 Sequence 11, Appli
12	503	10.3	113	4	US-09-635-872A-13 Sequence 13, Appli
13	479	9.8	502	4	US-09-636-077A-13 Sequence 13, Appli
14	479	9.8	502	4	US-09-636-060C-13 Sequence 13, Appli
15	479	9.8	502	4	US-09-986-552-13 Sequence 13, Appli
16	479	9.8	502	4	US-09-986-552-13 Sequence 13, Appli
17	255.5	5.4	545	4	US-08-936-107A-10 Sequence 10, Appli
18	255.5	5.4	545	4	US-08-936-107A-10 Sequence 10, Appli
19	240	4.9	364	4	US-09-328-352-5382 Sequence 5382, Ap
20	168.5	3.4	1015	3	US-09-113-825-1 Sequence 1, Appli
21	168.5	3.4	1015	3	US-09-113-825-1 Sequence 1, Appli
22	168.5	3.4	2471	1	US-08-185-432-16 Sequence 16, Appli
23	168.5	3.4	2471	1	US-08-185-432-16 Sequence 16, Appli
24	168.5	3.4	2471	1	US-08-532-384-19 Sequence 19, Appli
25	168.5	3.4	2471	1	US-08-532-384-19 Sequence 19, Appli
26	154.5	3.1	2321	4	US-09-230-652-2 Sequence 2, Appli
27	152	3.1	1964	4	US-09-467-997-1 Sequence 1, Appli

28	145.5	3.0	1068	1	US-08-537-210A-2 Sequence 2, Appli
29	145.5	3.0	1068	1	US-09-113-825-2 Sequence 2, Appli
30	145.5	3.0	2556	1	US-08-185-432-17 Sequence 17, Appli
31	145.5	3.0	2556	1	US-08-185-432-17 Sequence 17, Appli
32	145.5	3.0	2556	1	US-08-083-590A-20 Sequence 20, Appli
33	145.5	3.0	2556	1	US-08-532-384-20 Sequence 20, Appli
34	141	2.9	1078	1	US-08-264-534-32 Sequence 32, Appli
35	141	2.9	1078	1	US-08-083-590A-11 Sequence 11, Appli
36	141	2.9	1078	1	US-08-465-500-32 Sequence 32, Appli
37	141	2.9	1078	2	US-08-346-128-32 Sequence 32, Appli
38	141	2.9	1078	3	US-08-532-384-11 Sequence 11, Appli
39	141	2.9	1078	3	US-08-893-828-32 Sequence 32, Appli
40	138	2.8	1064	1	US-08-537-210A-3 Sequence 3, Appli
41	138	2.8	1064	3	US-09-113-825-3 Sequence 3, Appli
42	138	2.8	1064	3	US-08-185-432-18 Sequence 18, Appli
43	138	2.8	2523	4	US-08-899-232-3 Sequence 3, Appli
44	129	2.6	1139	1	US-08-537-210A-4 Sequence 4, Appli
45	129	2.6	1139	3	US-09-113-825-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-635-872A-1
Sequence 1, Application US/09635872A
Patent No. 6534300
GENERAL INFORMATION
APPLICANT: CANFIELD, WILLIAM
TITLE OR INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 195613US0
CURRENT APPLICATION NUMBER: US/09/635,872A
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 928
TYPE: PRT
ORGANISM: Homo sapiens
US-09-635-872A-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLFLLQRQTYTCLSHRYGLVYVVFVTVVAVAFQGEVYVLEWSDQYHVLFDSDYRDN	60
DB	1	MLFLLQRQTYTCLSHRYGLVYVVFVTVVAVAFQGEVYVLEWSDQYHVLFDSDYRDN	60
QY	61	AGKSFQNRCLPMPIDVYVTVWNGTDLDELKEQVREQWEQKAREILGNTEPTK	120
DB	61	AGKSFQNRCLPMPIDVYVTVWNGTDLDELKEQVREQWEQKAREILGNTEPTK	120
QY	121	KSEKQECCLTHCICVPMVLDPALPANTILKQVPSLYPSFHSASDIFNVAKRKNSTNV	180
DB	121	KSEKQECCLTHCICVPMVLDPALPANTILKQVPSLYPSFHSASDIFNVAKRKNSTNV	180
QY	181	SVVVFDSKTVEDAHSGSLKGNRQTVWRGYLTDDKQVPLVLMODLAFSGPPTFKET	240
DB	181	SVVVFDSKTVEDAHSGSLKGNRQTVWRGYLTDDKQVPLVLMODLAFSGPPTFKET	240
QY	241	NQTKTLPENLSSKVKLLQYSEASVALLKLNPKCFQELINKQTKNNMTIDGKELTISPA	300
DB	241	NQTKTLPENLSSKVKLLQYSEASVALLKLNPKCFQELINKQTKNNMTIDGKELTISPA	300
QY	301	YLLMDLSAISQKQEDDISASRFEDNELYSLSIRKHPVWRNFIIVNGQIPSWNL	360
DB	301	YLLMDLSAISQKQEDDISASRFEDNELYSLSIRKHPVWRNFIIVNGQIPSWNL	360
QY	361	DNPRVITVHQDYFRNLSHLPTESSPAIRBSHIREGLSKFTYLVNDVWFGQVWDDF	420
DB	361	DNPRVITVHQDYFRNLSHLPTESSPAIRBSHIREGLSKFTYLVNDVWFGQVWDDF	420

Db 361 DNRVATVTHQVFRNLSHLPTFSSPAISHHRIEGLSOKFIYINDVMFGKDYWPDF 420
QY 421 YSHSKQKXYLLTPVENCAGCGGSGSWIKDGYCDKACNNSACDMDGDDCGSGSGRYTAG 480
Db 421 YSHSKQKXYLLTPVENCAGCGGSGSWIKDGYCDKACNNSACDMDGDDCGSGSGRYTAG 480
QY 481 GGGTGSIGVGHFWQFGGGINSVSYCNQGCANSMADKFCDOACNVLSCGFDAGDGDGQHF 540
Db 481 GGGTGSIGVGHFWQFGGGINSVSYCNQGCANSMADKFCDOACNVLSCGFDAGDGDGQHF 540
QY 541 HELYKYLILPNQTHYIIPKGECLPYFSFAEVAKRGEGAYSNDPIIRHASIANKKYTHL 600
Db 541 HELYKYLILPNQTHYIIPKGECLPYFSFAEVAKRGEGAYSNDPIIRHASIANKKYTHL 600
QY 601 IMHSGNNAATTIHNLTFQNTNDEEFMOQITVEVDTRBPKLNSTAKGEMVLSPTLLP 660
Db 601 IMHSGNNAATTIHNLTFQNTNDEEFMOQITVEVDTRBPKLNSTAKGEMVLSPTLLP 660
QY 661 EAEILFEDIPEKRRPFKRGHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLOEH 720
Db 661 EAEILFEDIPEKRRPFKRGHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLOEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSOAKIKNOAIITDETNDSLVAPQEKVHKSILPNSLG 780
Db 721 GDTTLKGYNLSSKALLRSFLMNSOAKIKNOAIITDETNDSLVAPQEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKNGHDGONPDLLETTARFVETHTOKTIGGVTKERKPSLIY 840
Db 781 VSERLQRLTFPAVSVKNGHDGONPDLLETTARFVETHTOKTIGGVTKERKPSLIY 840
QY 841 PLESQWTKKKITGKEKENSMEENAENHIGTEVLGRKLOHYTDSYLGFLPWEKKYF 900
Db 841 PLESQWTKKKITGKEKENSMEENAENHIGTEVLGRKLOHYTDSYLGFLPWEKKYF 900
QY 901 QDLDEESBLKTQLAVFTDSKNTGRQK 928
Db 901 QDLDEESBLKTQLAVFTDSKNTGRQK 928

RESULT 2
US-09-636-077A-1
Sequence 1, Application US/09636077A
Patent No. 6537785
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM
TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
FILE REFERENCE: 195612USO
CURRENT APPLICATION NUMBER: US/09/636,077A
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.0
SEQ ID NO 1
LENGTH: 928
TYPE: PRT
ORGANISM: Homo sapiens
US-09-636-077A-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Claims to method of treatment

Db 121 KSEKQLECLLTHCICVPMVLDPALPANITLKDVPILYPSFHSASDIFNVAKKPNSTNY 180
QY 181 SVVFPDSTKVEDAHSGILKNGSRQTYWNGYLTITDEKVGVLVMDLAFSGFPPEFKET 240
Db 181 SVVFPDSTKVEDAHSGILKNGSRQTYWNGYLTITDEKVGVLVMDLAFSGFPPEFKET 240
QY 241 NQTKLPENLSKRYKLLQYSEASVALLKLNPKDFOELNQTKKNTIDGKELTISPA 300
Db 241 NQTKLPENLSKRYKLLQYSEASVALLKLNPKDFOELNQTKKNTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKODDISAREDENEELRYSLSRTERAPWNTFITNQIIPSWNL 360
Db 301 YLLMDLSAISQSKODDISAREDENEELRYSLSRTERAPWNTFITNQIIPSWNL 360
QY 361 DNRVATVTHQVFRNLSHLPTFSSPAISHHRIEGLSOKFIYINDVMFGKDYWPDF 420
Db 361 DNRVATVTHQVFRNLSHLPTFSSPAISHHRIEGLSOKFIYINDVMFGKDYWPDF 420
QY 421 YSHSKQKXYLLTPVENCAGCGGSGSWIKDGYCDKACNNSACDMDGDDCGSGSGRYTAG 480
Db 421 YSHSKQKXYLLTPVENCAGCGGSGSWIKDGYCDKACNNSACDMDGDDCGSGSGRYTAG 480
QY 481 GGGTGSIGVGHFWQFGGGINSVSYCNQGCANSMADKFCDOACNVLSCGFDAGDGDGQHF 540
Db 481 GGGTGSIGVGHFWQFGGGINSVSYCNQGCANSMADKFCDOACNVLSCGFDAGDGDGQHF 540
QY 541 HELYKYLILPNQTHYIIPKGECLPYFSFAEVAKRGEGAYSNDPIIRHASIANKKYTHL 600
Db 541 HELYKYLILPNQTHYIIPKGECLPYFSFAEVAKRGEGAYSNDPIIRHASIANKKYTHL 600
QY 601 IMHSGNNAATTIHNLTFQNTNDEEFMOQITVEVDTRBPKLNSTAKGEMVLSPTLLP 660
Db 601 IMHSGNNAATTIHNLTFQNTNDEEFMOQITVEVDTRBPKLNSTAKGEMVLSPTLLP 660
QY 661 EAEILFEDIPEKRRPFKRGHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLOEH 720
Db 661 EAEILFEDIPEKRRPFKRGHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLOEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSOAKIKNOAIITDETNDSLVAPQEKVHKSILPNSLG 780
Db 721 GDTTLKGYNLSSKALLRSFLMNSOAKIKNOAIITDETNDSLVAPQEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKNGHDGONPDLLETTARFVETHTOKTIGGVTKERKPSLIY 840
Db 781 VSERLQRLTFPAVSVKNGHDGONPDLLETTARFVETHTOKTIGGVTKERKPSLIY 840
QY 841 PLESQWTKKKITGKEKENSMEENAENHIGTEVLGRKLOHYTDSYLGFLPWEKKYF 900
Db 841 PLESQWTKKKITGKEKENSMEENAENHIGTEVLGRKLOHYTDSYLGFLPWEKKYF 900
QY 901 QDLDEESBLKTQLAVFTDSKNTGRQK 928
Db 901 QDLDEESBLKTQLAVFTDSKNTGRQK 928

RESULT 3
US-09-636-060C-1
Sequence 1, Application US/09636060C
Patent No. 6642038
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM M
TITLE OF INVENTION: GLCNAC PROSOPHOPRANSPERASE OF THE LYSOSOMAL TARGETING PATHWAY
FILE REFERENCE: 210119USOCONT
CURRENT APPLICATION NUMBER: US/09/636,060C
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.1
SEQ ID NO 1
LENGTH: 928
TYPE: PRT
ORGANISM: Homo sapiens

Claims to nucleic acid

US-09-636-060C-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTYTCLSHRGYVCFGLGVVTVTISAFOGFEVLEMSDQYHVLFDYSYRNI 60
DB 1 MLFKLQRTYTCLSHRGYVCFGLGVVTVTISAFOGFEVLEMSDQYHVLFDYSYRNI 60
QY 61 AGKSFQNRICLPMPIDVVYTWNGTDLLELKELOQVREOMEEOQKAMEIIGKNTTEPTK 120
DB 61 AGKSFQNRICLPMPIDVVYTWNGTDLLELKELOQVREOMEEOQKAMEIIGKNTTEPTK 120
QY 121 KSEKQECILTHCIVPMVLDPALPANITLKDVPSLYPSFASDIFNVAKRKNSTNV 180
DB 121 KSEKQECILTHCIVPMVLDPALPANITLKDVPSLYPSFASDIFNVAKRKNSTNV 180
QY 181 SVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLMQDLAFSLGPPPTFKET 240
DB 181 SVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLMQDLAFSLGPPPTFKET 240
QY 241 NQKTKLBNLSKRYKLOLYSEASVALLKKNPKDFOLNKQTKQNTIDGKELTISPA 300
DB 241 NQKTKLBNLSKRYKLOLYSEASVALLKKNPKDFOLNKQTKQNTIDGKELTISPA 300
QY 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRERHAPVVRNIFIVTNGQIPSWMLN 360
DB 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRERHAPVVRNIFIVTNGQIPSWMLN 360
QY 361 DNPRTVITVHODVFRNLSHLPTFSSPAIESHIREGLSOKFIYNDVDFGKVPDPDF 420
DB 361 DNPRTVITVHODVFRNLSHLPTFSSPAIESHIREGLSOKFIYNDVDFGKVPDPDF 420
QY 421 YSHSGQKRYLLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDMDGDCGSGSGSRYIAG 480
DB 421 YSHSGQKRYLLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDMDGDCGSGSGSRYIAG 480
QY 481 GGGTSGIGVHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCPDAGCCQDHF 540
DB 481 GGGTSGIGVHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCPDAGCCQDHF 540
QY 541 HELYKYLILPNQTHYIIPKGECLPYSPFAVAKRGVAGVSDNPIIRHASIANKWKTIHL 600
DB 541 HELYKYLILPNQTHYIIPKGECLPYSPFAVAKRGVAGVSDNPIIRHASIANKWKTIHL 600
QY 601 IMHSGNATTIHFNLTFOVNTDEEFKMOITVEVDITREGPKLNTAKGYENLVSPITLLP 660
DB 601 IMHSGNATTIHFNLTFOVNTDEEFKMOITVEVDITREGPKLNTAKGYENLVSPITLLP 660
QY 661 EAEIILFEDIPKEKRPFKRHVDNSTRAOEVEKILPVNISILPDAOALSTNTLDQLEH 720
DB 661 EAEIILFEDIPKEKRPFKRHVDNSTRAOEVEKILPVNISILPDAOALSTNTLDQLEH 720
QY 721 GGITLKGYVLSKALLRSFLMNSQAKIKNOAILITDETNDLSVAPOEQVHKSILPNSLG 780
DB 721 GGITLKGYVLSKALLRSFLMNSQAKIKNOAILITDETNDLSVAPOEQVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVGNDGONPDLLETTAFRVEYTHQTKTIGGNTVEKPSLIV 840
DB 781 VSERLQRLTFPAVSVKVGNDGONPDLLETTAFRVEYTHQTKTIGGNTVEKPSLIV 840
QY 841 PLESQMTXKKTITGKESKSMENNAENHIGTEVLLPRKLOHYTDSYLGFLPMEKKCYF 900
DB 841 PLESQMTXKKTITGKESKSMENNAENHIGTEVLLPRKLOHYTDSYLGFLPMEKKCYF 900
QY 901 QDLDEESLKTOLAYFTDSKNTGRQLK 928
DB 901 QDLDEESLKTOLAYFTDSKNTGRQLK 928

RESULT 4
US-09-986-552-1

; Sequence 1, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-552-1

Query Match 100.0%; Score 4907; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTYTCLSHRGYVCFGLGVVTVTISAFOGFEVLEMSDQYHVLFDYSYRNI 60
DB 1 MLFKLQRTYTCLSHRGYVCFGLGVVTVTISAFOGFEVLEMSDQYHVLFDYSYRNI 60
QY 61 AGKSFQNRICLPMPIDVVYTWNGTDLLELKELOQVREOMEEOQKAMEIIGKNTTEPTK 120
DB 61 AGKSFQNRICLPMPIDVVYTWNGTDLLELKELOQVREOMEEOQKAMEIIGKNTTEPTK 120
QY 121 KSEKQECILTHCIVPMVLDPALPANITLKDVPSLYPSFASDIFNVAKRKNSTNV 180
DB 121 KSEKQECILTHCIVPMVLDPALPANITLKDVPSLYPSFASDIFNVAKRKNSTNV 180
QY 181 SVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLMQDLAFSLGPPPTFKET 240
DB 181 SVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLMQDLAFSLGPPPTFKET 240
QY 241 NQKTKLBNLSKRYKLOLYSEASVALLKKNPKDFOLNKQTKQNTIDGKELTISPA 300
DB 241 NQKTKLBNLSKRYKLOLYSEASVALLKKNPKDFOLNKQTKQNTIDGKELTISPA 300
QY 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRERHAPVVRNIFIVTNGQIPSWMLN 360
DB 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRERHAPVVRNIFIVTNGQIPSWMLN 360
QY 361 DNPRTVITVHODVFRNLSHLPTFSSPAIESHIREGLSOKFIYNDVDFGKVPDPDF 420
DB 361 DNPRTVITVHODVFRNLSHLPTFSSPAIESHIREGLSOKFIYNDVDFGKVPDPDF 420
QY 421 YSHSGQKRYLLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDMDGDCGSGSGSRYIAG 480
DB 421 YSHSGQKRYLLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDMDGDCGSGSGSRYIAG 480
QY 481 GGGTSGIGVHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCPDAGCCQDHF 540
DB 481 GGGTSGIGVHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGCPDAGCCQDHF 540
QY 541 HELYKYLILPNQTHYIIPKGECLPYSPFAVAKRGVAGVSDNPIIRHASIANKWKTIHL 600
DB 541 HELYKYLILPNQTHYIIPKGECLPYSPFAVAKRGVAGVSDNPIIRHASIANKWKTIHL 600
QY 601 IMHSGNATTIHFNLTFOVNTDEEFKMOITVEVDITREGPKLNTAKGYENLVSPITLLP 660
DB 601 IMHSGNATTIHFNLTFOVNTDEEFKMOITVEVDITREGPKLNTAKGYENLVSPITLLP 660
QY 661 EAEIILFEDIPKEKRPFKRHVDNSTRAOEVEKILPVNISILPDAOALSTNTLDQLEH 720
DB 661 EAEIILFEDIPKEKRPFKRHVDNSTRAOEVEKILPVNISILPDAOALSTNTLDQLEH 720
QY 721 GGITLKGYVLSKALLRSFLMNSQAKIKNOAILITDETNDLSVAPOEQVHKSILPNSLG 780

Db 721 GDTLKGNYLSALRSFLMSQAKIKNOAIITDETNDLSVAQOEQVAKSLPNSLG 780
 QY 781 VSERLQRLTPPAVSVKNGHDGONPPLDLETTARFVETHTOKTIGANTKEKPSLIIV 840
 Db 781 VSERLQRLTPPAVSVKNGHDGONPPLDLETTARFVETHTOKTIGANTKEKPSLIIV 840
 QY 841 PLESQMTKEKKITGKEKENSMEENAHGIVTEVLGRKQHYTDSYLGFLPWEKKXYF 900
 Db 841 PLESQMTKEKKITGKEKENSMEENAHGIVTEVLGRKQHYTDSYLGFLPWEKKXYF 900
 QY 901 QDLDEESLKTQALAYFTDSKNTGRQK 928
 Db 901 QDLDEESLKTQALAYFTDSKNTGRQK 928

RESULT 5
 US-09-635-872A-15
 ; Sequence 15, Application US/09635872A
 ; Patent No. 6534300
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
 ; FILE REFERENCE: 195613050
 ; CURRENT APPLICATION NUMBER: US/09/635,872A
 ; CURRENT FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 15
 ; LENGTH: 908
 ; TYPE: PR1
 ; ORGANISM: Mus musculus
 US-09-635-872A-15

Query Match 79.4%; Score 3894; DB 4; Length 908;
 Best Local Similarity 79.9%; Pred. No. 0;
 Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MFKLQRTYCLSHRGVLCFVGAVVTIYSAFOGCVLWMSRDQHYLFDSDYRNI 60
 Db 1 MFKLQRTYCLSHRGVLCFVGAVVTIYSAFOGCVLWMSRDQHYLFDSDYRNI 60
 QY 61 AGSFQNRCLPMPIDVYVTVNGTDLBELKELQVREHMEBQKAMEIIGKNTTEPTK 120
 Db 61 AGSFQNRCLPMPIDVYVTVNGTDLBELKELQVREHMEBQKAMEIIGKNTTEPTK 120
 QY 121 KSEKQECCLTHCICIKVPMVLDPALPANITLKDVSLYPSFHSASDINNAKPKPSTNV 180
 Db 121 KSEKQECCLTHCICIKVPMVLDPALPANITLKDVSLYPSFHSASDINNAKPKPSTNV 180
 QY 181 SYVVPDSTKDVDAHSGLLKGNRSQRTWVRGYLTDDKEVPGVLMQDLAFLSGPPTFKET 240
 Db 181 SYVVPDSTKDVDAHSGLLKGNRSQRTWVRGYLTDDKEVPGVLMQDLAFLSGPPTFKET 240
 QY 241 NOLKTKLPEN-LSKYKLIQLYSEASVALLKLNKXDOELNKOTKMMTIDGKELTSP 299
 Db 241 NOLKTKLPEN-LSKYKLIQLYSEASVALLKLNKXDOELNKOTKMMTIDGKELTSP 299
 QY 300 AYLLWDLAISQSKDEDISARFEDNEELRYSLASIRRAHAWANIFIVTNGOIPSWLN 359
 Db 301 AYLLWDLAISQSKDEDISARFEDNEELRYSLASIRRAHAWANIFIVTNGOIPSWLN 360
 QY 360 LDNRPVTVTHODVVRANISHLPTFESSPAIESHIEHIEGLSQFIYLANDVWFGKDVPPD 419
 Db 361 LDNRPVTVTHODVVRANISHLPTFESSPAIESHIEHIEGLSQFIYLANDVWFGKDVPPD 420
 QY 420 FYSHSGQGVYLTWVPNCABGCPGSMWKDGYCDKACNNSACMDWGDGCSGNSGSR1A 479
 Db 421 FYSHSGQGVYLTWVPNCABGCPGSMWKDGYCDKACNNSACMDWGDGCSGNSGSR1A 480
 QY 480 GGGGTGSGVGHWPQGGGINSVYICNCGCANSWLADKFCQACNVLSCGFDAQDCGDH 539

Db 481 RGGGTGNGAGQHMQGGGINITISYCNCGCANSWLADKFCQACNVLSCGFDAQDCGDH 540
 QY 540 FHELYVYLLPNQTHYIIPKGECLPYFSAEVAKGVESAYDNPIIRASIANKXKTIH 599
 Db 541 FHELYVYLLPNQTHYVPRGEVLSYFSPANIAKRIEGTSDNPPIIRASIANKXKTIH 600
 QY 600 LINESGMATTHFNITFQNTNDEEFKQITEVDPREBPKINSTRQKXENYLSBITL 659
 Db 601 LIMPGENNATTYIFNLTQNADEBFKQIAEVDTRAPKINSTQKAYESLVSEVTP 660
 QY 660 PEAEIIFEDIPKEREPEKFRHDVNSTRAQSEVKIPLVNISLPEKQALSTINTDLOE 719
 Db 661 PQADVPEFVPEKREKRPKIRHADVATGRFOGEVKIPRNIISLPEKQALSTINTDLOE 720
 QY 720 HGDITLKGNYLSALRSFLMSQAKIKNOAIITDETNDLSVAQOEQVAKSLPNSL 779
 Db 721 RGDITLKGNYLSALRSFLMSQAKIKNOAIITDETNDLSVAQOEQVAKSLPNSL 777
 QY 780 GVSERLQRLTPPAVSVKNGHDGONPPLDLETTARFVETHTOKTIGANTKEKPSLI 839
 Db 778 AGHREBRTAPAEITVYGRDHALNPPVLETNRL-----AQPTLGVTKEKENSPLI 832
 QY 840 VPLESQMTKEKKITGKEKENSMEENAHGIVTEVLGRKQHYTDSYLGFLPWEKKXY 899
 Db 833 VPLESHLP-----KEESDRAEGNA--VVKELVGRRLQ--QNYPGFLPWEKKXY 879
 QY 900 QDLDEESLKTQALAYFTDSKNTGRQK 928
 Db 880 QDLDEESLKTQALAYFTDRKHTGRQK 908

RESULT 6
 US-09-636-077A-15
 ; Sequence 15, Application US/09636077A
 ; Patent No. 6537785
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
 ; FILE REFERENCE: 195612050
 ; CURRENT APPLICATION NUMBER: US/09/636,077A
 ; CURRENT FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 15
 ; LENGTH: 908
 ; TYPE: PR1
 ; ORGANISM: Mus musculus
 US-09-636-077A-15

Query Match 79.4%; Score 3894; DB 4; Length 908;
 Best Local Similarity 79.9%; Pred. No. 0;
 Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MFKLQRTYCLSHRGVLCFVGAVVTIYSAFOGCVLWMSRDQHYLFDSDYRNI 60
 Db 1 MFKLQRTYCLSHRGVLCFVGAVVTIYSAFOGCVLWMSRDQHYLFDSDYRNI 60
 QY 61 AGSFQNRCLPMPIDVYVTVNGTDLBELKELQVREHMEBQKAMEIIGKNTTEPTK 120
 Db 61 AGSFQNRCLPMPIDVYVTVNGTDLBELKELQVREHMEBQKAMEIIGKNTTEPTK 120
 QY 121 KSEKQECCLTHCICIKVPMVLDPALPANITLKDVSLYPSFHSASDINNAKPKPSTNV 180
 Db 121 KSEKQECCLTHCICIKVPMVLDPALPANITLKDVSLYPSFHSASDINNAKPKPSTNV 180
 QY 181 SYVVPDSTKDVDAHSGLLKGNRSQRTWVRGYLTDDKEVPGVLMQDLAFLSGPPTFKET 240
 Db 181 SYVVPDSTKDVDAHSGLLKGNRSQRTWVRGYLTDDKEVPGVLMQDLAFLSGPPTFKET 240
 QY 241 NOLKTKLPEN-LSKYKLIQLYSEASVALLKLNKXDOELNKOTKMMTIDGKELTSP 299

Db	241	SQUTKTLPRKAFPLKXIKLTLKLYSEASATLKLKNNPKGPFELNNQTKKNTITDKGELLTISF	300
Qy	300	AYLLMDLSAISQSDQEDDISASRPEDNEELRYSLSRSTERHAPWENIFIVTNGQIPSWLN	359
Db	301	AYLLMDLSAISQSDQEDDISASRPEDNEELRYSLSRSTERHAPWENIFIVTNGQIPSWLN	360
Qy	360	LDNERVATIVHODIFRNLISLPTFSSPALESHTREGLSQKRIYLANDVMEGKVPDPD	419
Db	361	LDNERVATIVHODIFQULSHLPTFSSPALESHTREGLSQKRIYLANDVMEGKVPDPD	420
Qy	420	FYHSKQKQKYYLLTPVPVNCACGCGPSYTKGQYCDKACNNSACDWDGDCGSGSGGRYTA	479
Db	421	FYHSKQKQKYYLLTPVPVNCACGCGPSYTKGQYCDKACNNSPCMDGNGCSGTAGRFFVA	480
Qy	480	GGGCTGSIIGVHPQFGGINSVYCNQGCANGLADKFCDDACNVLSGCPFADGCGDH	539
Db	481	RGSGTGNIGAGQHKQFQGGINTISYCNQGCANGLADKFCDDACNVLSGCPFADGCGDH	540
Qy	540	FHELYKYLILPNQTHYIIPKGECLPTPSFAEVAKQVEGAYSNDPIIRHASTANKKTIH	599
Db	541	FHELYKYLILPNQTHYVVPKGEYLSYFSPFANIARKIEGYSNDPIIRHASTANKKTIH	600
Qy	600	LIMSGNATTIHFNLFONTNDEEFPQITVEVDREGKLNSTAKQKXYNLSPITLL	659
Db	601	LIMSGNATTIYNNLQANDBERFIQJAVEVDTRAPKLNSTQKXYESLVSFVPL	660
Qy	660	PEALIEDIPKPKRPFKRHADVNSTRPAQEEVKIPLVNSILPDAQLSINTLDQLE	719
Db	661	POADVPEVDVPEKRPFKIRRHADVNTAGRQEEVKIPRVNISLLPKEACVRSNLDQLE	720
Qy	720	HGDITLGYNLSKSLRSPFNMSQAKINQOAITDENDSILVAQEQCVKXSLIPNLT	779
Db	721	RGDITLGYNLSKSLRSPFNLSLDTKIKPQA-RIDEIKGNLEVENPESHRR--PHGF	777
Qy	780	GVSRRLQRLTPPAVSVKVNGHDQGNPDLDETTAARFVETHQTKTIGGVTKERKPSLI	839
Db	778	AGEHSRRTWAPAEYTVKGRDHALNPPVLETNARL---AQPFLGTVSEKENLSPLI	832
Qy	840	VPLESNTKTEKLTGKEKENSMBEENAHNIGTEVILGRKLOHYHDSYLGFLPWEXKY	899
Db	833	VPPSHLP-----KEESDRAEGNA--VPKXELVPGRRLO--QNTPGFLPWEXKY	879
Qy	900	FQDILDEESIKQLAYFTDSKNTGRQLX	928
Db	880	FQDILDEESIKQLAYFTDRKHTGRQLX	908

RESULT 7

US-09-636-060C-15

; Sequence 15, Application US/09636060C

; Patent No. 6642038

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M

; TITLE OF INVENTION: GENUAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY

; FILE REFERENCE: 210119USOCONT

; CURRENT APPLICATION NUMBER: US/09/636,060C

; PRIOR FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 15

; LENGTH: 908

; TYPE: PRF

; ORGANISM: Mus musculus

US-09-636-060C-15

Query Match	79.4%	Score 3894	DB 4	Length 908
Best Local Similarity	79.9%	Pred. No. 0		
Matches 742	Conservative 61	Mismatches 104	Indels 22	Gaps 7

Db	1	MLFPLKQGYTTCLSHRGVLVYCFVGVVVTIVSAQFQBEVVLVEMSRDQYHVLFDSDYRDI	60
Qy	61	AGKSFONFLCLMPPIDVVYTVWNGTDELLKELQVREQMEBEQKAMREILCKNTTEPTK	120
Db	61	AGKSFONFLCLMPPIDVVYTVWNGTDELLKELQVREHMEBEQKAMREILCKNTTEPTK	120
Qy	121	KSEKQJECLLHFCIKVPMVLVDPALPANIITLKDVSLKPSFSPASADIFNVAKPKPSTNV	180
Db	121	KSEKQJECLLHFCIKVPMVLVDPALPANIITLKDLTLVPSFASDDMENVAKPKPSTNV	180
Qy	181	SVVVFDSFKDYEDAHSGLLKGNRSQTVWRGYLTTDKVEYVGLVMODLAFLSGFPPTFKET	240
Db	181	PVVVFEDTKDYEDAHAGFPKGGQQTDMVRAYLITTDKAPGLVLIQGLAFLSGFPPTFKET	240
Qy	241	NOILTKLPENLSSKVLLOLYSEASVALLKLNPKDQOLNKQTKKMTIDGKLLTSP	299
Db	241	SOILTKLPKPAFPEILKILRLYSEASVALLKLNPKGQOLNKQTKKMTIDGKLLTSP	300
Qy	300	AYLMDLSAISQSKODEDISASRFEDNEELRSLRSIRRHAPWANIETVINGQJPSMLN	359
Db	301	AYLMDLSAISQSKODEDASASRFEDNEELRSLRSIRRHAPWANIETVINGQJPSMLN	360
Qy	360	LDNPRVTVITVHQDVFRNLSHLPTFESSPAIESHIRIEGLSQFYIYLANDVMFKDVPDD	419
Db	361	LDNPRVTVITVHQDIFQNLSHLPTFESSPAIESHIRIEGLSQFYIYLANDVMFKDVPDD	420
Qy	420	FYSHSKQGXVLYVMPVPCAGCGPGSWIKDGCDDAKCNNSACDDMGCGGSGSGSRYIA	479
Db	421	FYSHSKGKXVLYVMPVPCAGCGPGSWIKDGCDDACNTSPCDMGGGNSGTAAGRFA	480
Qy	480	GGGGTGSIGVHPWQFGGGINSVSYCNOGCANSWLADKFCDDACNVLSGCFPAGCGGODH	539
Db	481	RGGGTGNIGAGQHMGFGGGINITISYCNOGCANSWLADKFCDDACNVLSGCFPAGCGGODH	540
Qy	540	FHELYKVLLENOYHYIIPKGECLPYESFPAEYAKGCVGANSDNPIIRHASJANKWKTH	599
Db	541	FHELYKVLLENOYHYVVPKGEYLSYFSPANIJARRIGTYSDNPIIRHASJANKWKTH	600
Qy	600	LIMSGMNAATIHENLTFONTNDEBFKQITVEVDTREGPKLNSPAAQKYEVLVSPITLL	659
Db	601	LIMSGMNAATIHENLTIQANDEBFKQIYAEVDTTRAPKLNSTYQARAYESLVPTLL	660
Qy	660	PEAEILFEDLPEKRRPFKRDVNSTRAQOEVKIPLVNISLBPKAQSLINTLDLOIE	719
Db	661	FOADVPFEDVPEKEKFPKIRRHADVATGRFOEYVKIPRVNISLBPKEAQRVRLSNDLOIE	720
Qy	720	HGDTTLKYNLSKSLILBSFLMNSQKAKITKQAITTDETNISLVAPQEKQYVKSILPNL	779
Db	721	RGDITTLKYNLSKSLILBSFLGNSLDTKIKQOA-RTDITKGLLEVQDPSHRR--PHGF	777
Qy	780	GVSERLQGLTFPAVSVKNGHDQGNPDLDETTARFVEVETQKTIIGANVYKEKPSLI	839
Db	778	AGEHRSSEWTPAELVTVYVKG RDHALNP RPVLETNARL-----AQPTLGVTVSKENLSPLI	832
Qy	840	VPLBSQMTKEKKITGKEKENSRLMEENAHNIIVTEVLLGRKLOHYTDSYLGLVPEKKKY	899
Db	833	VPESHLP-----KEEESDRABGNA--VPVKELVGRRIQ--QNYPGFLPWEKKKY	879
Qy	900	FODLLDEBSLKTOLAYFTDSKNTGRQK	928
Db	880	FODLLDEBSLKTOLAYFTDRKHTGRQK	908

```

RESULT 8
US-09-986-552-15
; Sequence 15, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 2150890577DIv
; CURRENT APPLICATION NUMBER: US/09/986,552

```

; CURRENT FILING DATE: 2001-11-09
 ; PRIOR APPLICATION NUMBER: 09/635,872
 ; PRIOR FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; PRIOR FILING DATE: 1999-09-14
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 15
 ; LENGTH: 908
 ; TYPE: PR1
 ; ORGANISM: Mus musculus
 ; US-09-986-552-15

Query Match 79.4%; Score 3894; DB 4; Length 908;
 Best Local Similarity 79.9%; Pred. No. 0;
 Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

```

QY 1 MFKLLQRTYTCLSHRYGLVCPGVVTVTISAFQFGVGVLEMSRDQYHLFDSYRNI 60
DB 1 MFKLLQRTYTCLSHRYGLVCPGVVTVTISAFQFGVGVLEMSRDQYHLFDSYRNI 60
QY 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELQOVREMEBEQKAMEIIGKTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYVTVWNGTDLLELQOVREMEBEQKAMEIIGKTEPTK 120
QY 121 KSEKQECLETHCIVKPMVLDPALPANTLKDVSLVPSFHSASDIFVAKKPKSTNV 180
DB 121 KSEKQECLETHCIVKPMVLDPALPANTLKDVSLVPSFHSASDIFVAKKPKSTNV 180
QY 181 SYVPEPSTKDVEDASGLKGNRSQTVWRGVLTTDKREYVGLVMDLFLSGFPTEKET 240
DB 181 SYVPEPSTKDVEDASGLKGNRSQTVWRGVLTTDKREYVGLVMDLFLSGFPTEKET 240
QY 241 NQKTLQTPRN-ISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDGKELTISP 299
DB 241 NQKTLQTPRN-ISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDGKELTISP 299
QY 300 ATLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRHAPWVNIPIVINGQIPSWLN 359
DB 300 ATLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRHAPWVNIPIVINGQIPSWLN 359
QY 360 LNPRTVITVHODVPRNLSHLPTFSSPAIESHIREGLSOKETIYLANDVYFGKDVPPD 419
DB 360 LNPRTVITVHODVPRNLSHLPTFSSPAIESHIREGLSOKETIYLANDVYFGKDVPPD 419
QY 420 FYSHSGQKQVYLTWPVPCNCAEGCPGSMWKDGYCDKACNNSACDMGDGSCGSGGSRVIA 479
DB 420 FYSHSGQKQVYLTWPVPCNCAEGCPGSMWKDGYCDKACNNSACDMGDGSCGSGGSRVIA 479
QY 480 GGGGTGSIQVGHWPQFGGGINSVSYCNOGCANSWLADKFCQACNVLSGCFDAGDCGDH 539
DB 480 GGGGTGSIQVGHWPQFGGGINSVSYCNOGCANSWLADKFCQACNVLSGCFDAGDCGDH 539
QY 540 FHELYKVTLLPNOTYIIPKGECLPYFSEFABYAKGAVGANSNPIITHASIAKWKTIH 599
DB 540 FHELYKVTLLPNOTYIIPKGECLPYFSEFABYAKGAVGANSNPIITHASIAKWKTIH 599
QY 541 FHELYKVTLLPNOTYIIPKGECLPYFSEFABYAKGAVGANSNPIITHASIAKWKTIH 600
DB 541 FHELYKVTLLPNOTYIIPKGECLPYFSEFABYAKGAVGANSNPIITHASIAKWKTIH 600
QY 600 LTHSGMNAATTIHFNLTFQNTNDEEFKQOITVEVDTRBGPKLNTAQKGVENLVSPTLL 659
DB 600 LTHSGMNAATTIHFNLTFQNTNDEEFKQOITVEVDTRBGPKLNTAQKGVENLVSPTLL 659
QY 660 PEAELIFEDIPEKKEFPKPKRDVNSTRAQBEVKIPLVNIISLPLKQJLSLNTLDLOLE 719
DB 660 PEAELIFEDIPEKKEFPKPKRDVNSTRAQBEVKIPLVNIISLPLKQJLSLNTLDLOLE 719
QY 720 HODITLKGNLSKSLNSFLMNSOHAQIKNOAITDDETNSLVAPOKQVHKISILPNSL 779
DB 720 HODITLKGNLSKSLNSFLMNSOHAQIKNOAITDDETNSLVAPOKQVHKISILPNSL 779
QY 780 GVSESLQRLTPPAVSKVNGHDQGNPPDLSTTARFVEVHTQKTCIGAVTKKPKPSLI 839
DB 780 GVSESLQRLTPPAVSKVNGHDQGNPPDLSTTARFVEVHTQKTCIGAVTKKPKPSLI 839
QY 778 AGEHSERWTAAETVYKGRDHALNPPVLETNARL-----AQTLLGVTVSKENLSPLI 832
DB 778 AGEHSERWTAAETVYKGRDHALNPPVLETNARL-----AQTLLGVTVSKENLSPLI 832
  
```

```

QY 840 VPLESQMTKEKKITGKEKENSEMEANAENHIGTEVLLGRKLOHYTDSYLGFLPWKCKY 899
DB 840 VPLESQMTKEKKITGKEKENSEMEANAENHIGTEVLLGRKLOHYTDSYLGFLPWKCKY 899
QY 833 VPPESHLP-----KEEEDRAEGNA---VPVKELVPRRLQ---QNPFGFLPWKCKY 879
DB 833 VPPESHLP-----KEEEDRAEGNA---VPVKELVPRRLQ---QNPFGFLPWKCKY 879
QY 900 FODLDBEESTKQOLAYFTDSKTRGOLK 928
DB 900 FODLDBEESTKQOLAYFTDSKTRGOLK 928
QY 880 FODLDBEESTKQOLAYFTDRKHTGOLK 908
DB 880 FODLDBEESTKQOLAYFTDRKHTGOLK 908
  
```

RESULT 9
 US-09-635-872A-11
 ; Sequence 11; Application US/09635872A
 ; Patent No. 6534300
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
 ; FILE REFERENCE: 195612050
 ; CURRENT APPLICATION NUMBER: US/09/635,872A
 ; CURRENT FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 11
 ; LENGTH: 113
 ; TYPE: PR1
 ; ORGANISM: Rattus rattus
 ; US-09-635-872A-11

Query Match 10.3%; Score 503; DB 4; Length 113;
 Best Local Similarity 92.5%; Pred. No. 1,1e-36;
 Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

QY 233 FPPFTFETNOLKTKLPENISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDG 292
DB 233 FPPFTFETNOLKTKLPENISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDG 292
QY 1 FPPFTFETNOLKTKLPENISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDG 60
DB 1 FPPFTFETNOLKTKLPENISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDG 60
QY 293 KELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRH 339
DB 293 KELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRH 339
QY 61 KELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRH 107
DB 61 KELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRH 107
  
```

RESULT 10
 US-09-636-077A-11
 ; Sequence 11; Application US/09636077A
 ; Patent No. 6537765
 ; GENERAL INFORMATION:
 ; APPLICANT: CANFIELD, WILLIAM
 ; TITLE OF INVENTION: METHODS OF TREATING LYOSOMAL STORAGE DISEASE
 ; FILE REFERENCE: 195612050
 ; CURRENT APPLICATION NUMBER: US/09/636,077A
 ; CURRENT FILING DATE: 2000-08-10
 ; PRIOR APPLICATION NUMBER: 60/153,831
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 11
 ; LENGTH: 113
 ; TYPE: PR1
 ; ORGANISM: Rattus rattus
 ; US-09-636-077A-11

Query Match 10.3%; Score 503; DB 4; Length 113;
 Best Local Similarity 92.5%; Pred. No. 1,1e-36;
 Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

QY 233 FPPFTFETNOLKTKLPENISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDG 292
DB 233 FPPFTFETNOLKTKLPENISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDG 292
QY 1 FPPFTFETNOLKTKLPENISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDG 60
DB 1 FPPFTFETNOLKTKLPENISSKVKLIQLYSEASVALIKLNPKDQOELNKQTKKMTIDG 60
QY 293 KELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRH 339
DB 293 KELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRH 339
QY 61 KELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRH 107
DB 61 KELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRH 107
  
```


RESULT 11
US-09-636-060C-11
; Sequence 11, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLUTAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119050CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRF
; ORGANISM: Rattus rattus
US-09-636-060C-11

Query Match 10.3%; Score 503; DB 4; Length 113;
Best Local Similarity 92.5%; Pred. No. 1.1e-36;
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 233 PEPFRETQKTKLPENLSSKVKLLQYSEASVALLKLNPKDQELNKQTKKMTIDG 292
DB 1 PEPFRETQKTKLPENLSSKVKLLQYSEASVALLKLNPKDQELNKQTKKMTIDG 60
QY 293 KELTISPAYLWDLAISOSKODEDISASRFEDNEELRYSLSIERH 339
DB 61 KELTISPAYLWDLAISOSKODEDVASASRFEDNEELRYSLSIERH 107

RESULT 12
US-09-986-552-11
; Sequence 11, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 2150890577DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRF
; ORGANISM: Rattus rattus
US-09-986-552-11

Query Match 10.3%; Score 503; DB 4; Length 113;
Best Local Similarity 92.5%; Pred. No. 1.1e-36;
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 233 PEPFRETQKTKLPENLSSKVKLLQYSEASVALLKLNPKDQELNKQTKKMTIDG 292
DB 1 PEPFRETQKTKLPENLSSKVKLLQYSEASVALLKLNPKDQELNKQTKKMTIDG 60
QY 293 KELTISPAYLWDLAISOSKODEDISASRFEDNEELRYSLSIERH 339
DB 61 KELTISPAYLWDLAISOSKODEDVASASRFEDNEELRYSLSIERH 107

RESULT 13
US-09-635-872A-13
; Sequence 13, Application US/09635872A

; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613050
; CURRENT APPLICATION NUMBER: US/09/635,872A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRF
; ORGANISM: Drosophila melanogaster
US-09-635-872A-13

Query Match 9.8%; Score 479; DB 4; Length 502;
Best Local Similarity 55.4%; Pred. No. 1.8e-33;
Matches 82; Conservative 29; Mismatches 37; Indels 0; Gaps 0;

QY 322 RFEDNEELRYSLSIERHAPVBNFIYVINGQIPSWLNDNPRVTIVTHQDVFRNLSDP 381
DB 4 RFDDXNLEIRYSLSIERHAPVBNFIYVINGQIPSWLNDNPRVTIVTHQDVFRNLSDP 63
QY 382 TFSSPAIESHIRIEGISOKEIYLANDVMPGKDVDPDYSKSGQKYLTPVPNCAG 441
DB 64 TFSSPAIESHIRIEGISOKEIYLANDVMPGKDVDPDYSKSGQKYLTPVPNCAG 123
QY 442 CPGSWKDGYCDKACNNNSACDWDGDCS 469
DB 124 CPWTYIGDAGCDRHCNDACQDFGDCS 151

RESULT 14
US-09-636-077A-13
; Sequence 13, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612050
; CURRENT APPLICATION NUMBER: US/09/636,077A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRF
; ORGANISM: Drosophila melanogaster
US-09-636-077A-13

Query Match 9.8%; Score 479; DB 4; Length 502;
Best Local Similarity 55.4%; Pred. No. 1.8e-33;
Matches 82; Conservative 29; Mismatches 37; Indels 0; Gaps 0;

QY 322 RFEDNEELRYSLSIERHAPVBNFIYVINGQIPSWLNDNPRVTIVTHQDVFRNLSDP 381
DB 4 RFDDXNLEIRYSLSIERHAPVBNFIYVINGQIPSWLNDNPRVTIVTHQDVFRNLSDP 63
QY 382 TFSSPAIESHIRIEGISOKEIYLANDVMPGKDVDPDYSKSGQKYLTPVPNCAG 441
DB 64 TFSSPAIESHIRIEGISOKEIYLANDVMPGKDVDPDYSKSGQKYLTPVPNCAG 123
QY 442 CPGSWKDGYCDKACNNNSACDWDGDCS 469
DB 124 CPWTYIGDAGCDRHCNDACQDFGDCS 151

RESULT 15
US-09-636-060C-13

```

; Sequence 13, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLUCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119050CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
US-09-636-060C-13

```

```

Query Match          9.8%; Score 479; DB 4; Length 502;
Best Local Similarity 55.4%; Pred. No. 1.8e-33;
Matches 82; Conservative 29; Mismatches 37; Indels 0; Gaps 0;

```

```

QY 322 REEDNEELRYSLRSIERHAPWVRNFIYVNGQIPSMNDNPRVTIYTHQDVFRNLSHP 381
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 4  RPDCKNEELRYSLRSLEKHAMIRHYIYVNGQIPSMNDLSYERYIVPEHVLAPDPQLP 63
QY 382 TFSSPAIESHRIEGLSQKFIYLANDVWFGKQWPDDEYSHSKGQKYYLTWPVNCAG 441
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 64 TFSSSAIEFTLHRIPTLSRFLYLNDDIFLGAPLYPEDELYTEAGVRYQAMVPGCALD 123
QY 442 CPGSWTKDGYCDKACNNSACDWDGDCS 469
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 124 CPWTYIGDQACDRHCNIDACQPDGDCS 151

```

Search completed: July 26, 2004, 11:13:18
 Job time : 28.1576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:11:44 ; Search time 71.3389 Seconds

(without alignments)
4073.614 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907
Sequence: 1 MLFKLQRQYITCLSHRYGL.....SLKTQLAVFTDSKNTGRQLK 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4907	100.0	928	9	US-09-895-072-1
2	4907	100.0	928	9	US-09-895-552-1
3	4907	100.0	928	14	US-10-023-888-4
4	4907	100.0	928	14	US-10-023-889-4
5	4907	100.0	928	14	US-10-023-889-4
6	4907	100.0	928	14	US-10-023-890-4
7	4907	100.0	928	14	US-10-024-197-4
8	4907	100.0	928	14	US-10-023-894-4
9	4907	100.0	928	14	US-10-306-686-1
10	4907	100.0	928	14	US-10-023-888-2
11	4907	100.0	928	14	US-10-023-889-2
12	4907	100.0	928	14	US-10-023-890-2
13	4907	100.0	928	14	US-10-024-197-2
14	4907	100.0	928	14	US-10-023-894-2
15	4907	100.0	928	14	US-10-120-801-88

16	4391	89.5	1459	15	US-10-120-801-26	Sequence 26, Appl
17	3894	79.4	908	9	US-09-895-072-15	Sequence 15, Appl
18	3894	79.4	908	9	US-09-895-552-15	Sequence 15, Appl
19	3894	79.4	908	14	US-10-023-888-9	Sequence 9, Appl
20	3894	79.4	908	14	US-10-023-888-9	Sequence 9, Appl
21	3894	79.4	908	14	US-10-023-889-9	Sequence 9, Appl
22	3894	79.4	908	14	US-10-024-197-9	Sequence 9, Appl
23	3894	79.4	908	14	US-10-023-894-9	Sequence 9, Appl
24	3894	79.4	908	14	US-10-306-686-15	Sequence 15, Appl
25	2021	41.2	663	15	US-10-120-801-87	Sequence 87, Appl
26	1900	38.7	367	9	US-09-864-761-42893	Sequence 42893, A
27	1900	38.7	367	14	US-10-023-386-32927	Sequence 32927, A
28	1292	26.3	242	15	US-10-139-794-173	Sequence 173, App
29	734.5	15.0	384	15	US-10-120-801-89	Sequence 89, Appl
30	503	10.3	113	9	US-09-895-072-11	Sequence 11, Appl
31	503	10.3	113	9	US-09-895-552-11	Sequence 11, Appl
32	503	10.3	113	14	US-10-023-888-14	Sequence 14, Appl
33	503	10.3	113	14	US-10-023-889-14	Sequence 14, Appl
34	503	10.3	113	14	US-10-023-890-14	Sequence 14, Appl
35	503	10.3	113	14	US-10-024-197-14	Sequence 14, Appl
36	503	10.3	113	14	US-10-023-894-14	Sequence 14, Appl
37	503	10.3	113	14	US-10-306-686-11	Sequence 11, Appl
38	488	9.9	652	15	US-10-120-801-91	Sequence 91, Appl
39	479	9.8	502	9	US-09-895-072-13	Sequence 13, Appl
40	479	9.8	502	9	US-09-895-552-13	Sequence 13, Appl
41	479	9.8	502	14	US-10-023-888-16	Sequence 16, Appl
42	479	9.8	502	14	US-10-023-889-16	Sequence 16, Appl
43	479	9.8	502	14	US-10-023-890-16	Sequence 16, Appl
44	479	9.8	502	14	US-10-024-197-16	Sequence 16, Appl
45	479	9.8	502	14	US-10-023-894-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-895-072-1
Sequence 1, Application US/09895072
Patent No. US2002002550A1
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM M
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 210119US00CONT
CURRENT APPLICATION NUMBER: US/09/895,072
CURRENT FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
PRIOR APPLICATION NUMBER: US 09/635,872
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 928
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-072-1

Query Match 100.0% Score 4907; DB 9; Length 928;
Best Local Similarity 100.0%; Pred No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRQYITCLSHRYGLVYVVIYSAPFGFGEVLEWMSRDQYHVLFDSDYRNI 60
1 MLFKLQRQYITCLSHRYGLVYVVIYSAPFGFGEVLEWMSRDQYHVLFDSDYRNI 60
Db 1 MLFKLQRQYITCLSHRYGLVYVVIYSAPFGFGEVLEWMSRDQYHVLFDSDYRNI 60

QY 61 AGKSPNRICLPMPIDIVYVTVNGTDLELKEIQVRQEMERQKMRILIGNTEPEPK 120
61 AGKSPNRICLPMPIDIVYVTVNGTDLELKEIQVRQEMERQKMRILIGNTEPEPK 120
Db 61 AGKSPNRICLPMPIDIVYVTVNGTDLELKEIQVRQEMERQKMRILIGNTEPEPK 120

QY 121 KSEKQECILTHCICKVPMVLDPALPANTILKDVPSLYSFHSASDIFVNAKPKNSTNV 180
121 KSEKQECILTHCICKVPMVLDPALPANTILKDVPSLYSFHSASDIFVNAKPKNSTNV 180
Db 121 KSEKQECILTHCICKVPMVLDPALPANTILKDVPSLYSFHSASDIFVNAKPKNSTNV 180

QY 181 SVVFPDSTKDVEDASGLKNGSRQTVWEGVLTIDKEVGLVLMODLAFLSGFPPTFKET 240
DB 181 SVVFPDSTKDVEDASGLKNGSRQTVWEGVLTIDKEVGLVLMODLAFLSGFPPTFKET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFOLNKQTKKNTIDGKELTISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFOLNKQTKKNTIDGKELTISPA 300
QY 301 YLLMDLSAISQKQEDDISASRFEDNEELRYSLSIESHAPVWRNIFVTNGQIIPSWML 360
DB 301 YLLMDLSAISQKQEDDISASRFEDNEELRYSLSIESHAPVWRNIFVTNGQIIPSWML 360
QY 361 DNPRTVITVHODVFNLSHLPTFSSPAIESHHRLEGSKQFIYLANDVMGKVPWPDF 420
DB 361 DNPRTVITVHODVFNLSHLPTFSSPAIESHHRLEGSKQFIYLANDVMGKVPWPDF 420
QY 421 YHSHKQKQKYLITWVPNCAEGCPGSMIKDGYCDKACNNACDMDGDCSGNSGSRITAG 480
DB 421 YHSHKQKQKYLITWVPNCAEGCPGSMIKDGYCDKACNNACDMDGDCSGNSGSRITAG 480
QY 481 GGGTGSIGVHPWQGGGINSVSYCNOGCANSLADKFCDDACNTLSCGPDAGDGGQDHF 540
DB 481 GGGTGSIGVHPWQGGGINSVSYCNOGCANSLADKFCDDACNTLSCGPDAGDGGQDHF 540
QY 541 HELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSNDPIIRHASIANKWKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSNDPIIRHASIANKWKTIHL 600
QY 601 IYHSGNNATTHFNLTFOQNTDEEFKMOIYEVDRREGPKLNSTAKQYENLVSPITLLP 660
DB 601 IYHSGNNATTHFNLTFOQNTDEEFKMOIYEVDRREGPKLNSTAKQYENLVSPITLLP 660
QY 661 EAEIILFEDIPEKRPKPKFRHDVNSTRAQEBVKIPLVNISSLPRDQSLNTLIDLOEH 720
DB 661 EAEIILFEDIPEKRPKPKFRHDVNSTRAQEBVKIPLVNISSLPRDQSLNTLIDLOEH 720
QY 721 GDTTLKGYNLSSALLRSFLMNSQAKIKNOAIITDETNDSLVAPQEKQVHKSILPNSLG 780
DB 721 GDTTLKGYNLSSALLRSFLMNSQAKIKNOAIITDETNDSLVAPQEKQVHKSILPNSLG 780
QY 781 VSEIRLQRLTFPAVSXKVNHDGQNPDLDETTAFREYETHQKTIIGAVTKEKPPSLIV 840
DB 781 VSEIRLQRLTFPAVSXKVNHDGQNPDLDETTAFREYETHQKTIIGAVTKEKPPSLIV 840
QY 841 PLESQWTKKTKITGKEKENSMEENAEHNIIVTEVLGRKLOHYTDSYLGFLPWEKCKYF 900
DB 841 PLESQWTKKTKITGKEKENSMEENAEHNIIVTEVLGRKLOHYTDSYLGFLPWEKCKYF 900
QY 901 QDLDBEESLKTQLAYFTDSKNTGRQK 928
DB 901 QDLDBEESLKTQLAYFTDSKNTGRQK 928

RESULT 2

US-09-986-552-1
Sequence 1, Application US/09986552
Patent No. US20020150981A1
GENERAL INFORMATION:
APPLICANT: CANFIELD, William
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 215089US77DIIV
CURRENT APPLICATION NUMBER: US/09/986,552
CURRENT FILING DATE: 2001-11-09
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 928
TYPE: PRF
ORGANISM: Homo sapiens

US-09-986-552-1
Query Match 100.0%; Score 4907; DB 9; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEFKLIQROTYYTCLSRIGLYCYFLGAVTYYISARQFGVYLENSRDQYHVLFDSDYRNI 60
DB 1 MEFKLIQROTYYTCLSRIGLYCYFLGAVTYYISARQFGVYLENSRDQYHVLFDSDYRNI 60
QY 61 AGSFQNRCLPMPIDVYVTVWNGTDLLELLEKLOQREBMEBQKAMEIILGKNTTEPTK 120
DB 61 AGSFQNRCLPMPIDVYVTVWNGTDLLELLEKLOQREBMEBQKAMEIILGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKYPMLVDPALPANTTLKDVSLYPSFHSASDIPNAKPNSTNV 180
DB 121 KSEKQLECLTHCIKYPMLVDPALPANTTLKDVSLYPSFHSASDIPNAKPNSTNV 180
QY 181 SVVFPDSTKDVEDASGLKNGSRQTVWEGVLTIDKEVGLVLMODLAFLSGFPPTFKET 240
DB 181 SVVFPDSTKDVEDASGLKNGSRQTVWEGVLTIDKEVGLVLMODLAFLSGFPPTFKET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFOLNKQTKKNTIDGKELTISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFOLNKQTKKNTIDGKELTISPA 300
QY 301 YLLMDLSAISQKQEDDISASRFEDNEELRYSLSIESHAPVWRNIFVTNGQIIPSWML 360
DB 301 YLLMDLSAISQKQEDDISASRFEDNEELRYSLSIESHAPVWRNIFVTNGQIIPSWML 360
QY 361 DNPRTVITVHODVFNLSHLPTFSSPAIESHHRLEGSKQFIYLANDVMGKVPWPDF 420
DB 361 DNPRTVITVHODVFNLSHLPTFSSPAIESHHRLEGSKQFIYLANDVMGKVPWPDF 420
QY 421 YHSHKQKQKYLITWVPNCAEGCPGSMIKDGYCDKACNNACDMDGDCSGNSGSRITAG 480
DB 421 YHSHKQKQKYLITWVPNCAEGCPGSMIKDGYCDKACNNACDMDGDCSGNSGSRITAG 480
QY 481 GGGTGSIGVHPWQGGGINSVSYCNOGCANSLADKFCDDACNTLSCGPDAGDGGQDHF 540
DB 481 GGGTGSIGVHPWQGGGINSVSYCNOGCANSLADKFCDDACNTLSCGPDAGDGGQDHF 540
QY 541 HELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSNDPIIRHASIANKWKTIHL 600
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSNDPIIRHASIANKWKTIHL 600
QY 601 IYHSGNNATTHFNLTFOQNTDEEFKMOIYEVDRREGPKLNSTAKQYENLVSPITLLP 660
DB 601 IYHSGNNATTHFNLTFOQNTDEEFKMOIYEVDRREGPKLNSTAKQYENLVSPITLLP 660
QY 661 EAEIILFEDIPEKRPKPKFRHDVNSTRAQEBVKIPLVNISSLPRDQSLNTLIDLOEH 720
DB 661 EAEIILFEDIPEKRPKPKFRHDVNSTRAQEBVKIPLVNISSLPRDQSLNTLIDLOEH 720
QY 721 GDTTLKGYNLSSALLRSFLMNSQAKIKNOAIITDETNDSLVAPQEKQVHKSILPNSLG 780
DB 721 GDTTLKGYNLSSALLRSFLMNSQAKIKNOAIITDETNDSLVAPQEKQVHKSILPNSLG 780
QY 781 VSEIRLQRLTFPAVSXKVNHDGQNPDLDETTAFREYETHQKTIIGAVTKEKPPSLIV 840
DB 781 VSEIRLQRLTFPAVSXKVNHDGQNPDLDETTAFREYETHQKTIIGAVTKEKPPSLIV 840
QY 841 PLESQWTKKTKITGKEKENSMEENAEHNIIVTEVLGRKLOHYTDSYLGFLPWEKCKYF 900
DB 841 PLESQWTKKTKITGKEKENSMEENAEHNIIVTEVLGRKLOHYTDSYLGFLPWEKCKYF 900
QY 901 QDLDBEESLKTQLAYFTDSKNTGRQK 928
DB 901 QDLDBEESLKTQLAYFTDSKNTGRQK 928

RESULT 3
US-10-023-888-4

```
/ Sequence 4, Application US/10023888
/ Publication No. US20030119088A1
/ GENERAL INFORMATION:
/ APPLICANT: CANFIELD, William
/ TITLE OF INVENTION: SOLUBLE GUNAC PHOSPHOTRANSFERASE
/ FILE REFERENCE: 203515US77
/ CURRENT APPLICATION NUMBER: US/10/023, 888
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 4
/ LENGTH: 928
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-10-023-888-4
```

```
Query Match 100.0%; Score 4907; DB 14; Length 928;
```

```
Best Local Similarity 100.0%; Pred. No. 0;
```

```
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MFKLORQTYTCLSHRYGLVYCFGLGVVYTVISAFQFGEVYLEWSDQYHVLFDSTYRDN 60
DB 1 MFKLORQTYTCLSHRYGLVYCFGLGVVYTVISAFQFGEVYLEWSDQYHVLFDSTYRDN 60
QY 61 AGKSFQNRCLPMPIDVYTVWNGTDLLELKELOQVREQMEBEQKAMREILGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYTVWNGTDLLELKELOQVREQMEBEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKYPMLVLDPALPANITLKDVPSLYPSFHSASDIPVAKKNSSTNV 180
DB 121 KSEKQLECLTHCIKYPMLVLDPALPANITLKDVPSLYPSFHSASDIPVAKKNSSTNV 180
QY 181 SYVVEPSTKDVDAHSGLLKGNRSQTVWNGYLTDDKEVGLVLMODLAFSGPPTFKET 240
DB 181 SYVVEPSTKDVDAHSGLLKGNRSQTVWNGYLTDDKEVGLVLMODLAFSGPPTFKET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFOELNKQTKKNTIDGKEITISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFOELNKQTKKNTIDGKEITISPA 300
QY 301 YLLMDLSAISQSKODEDISASRFEDNEBELRYSLSIERHAPVWRNIFIVTNGQIPSWNL 360
DB 301 YLLMDLSAISQSKODEDISASRFEDNEBELRYSLSIERHAPVWRNIFIVTNGQIPSWNL 360
QY 361 DNPRTVITVHQDVFNRLSHLPTSSPAIESHIRIEGLSQKFIYLNDDVMFGKDWPPDF 420
DB 361 DNPRTVITVHQDVFNRLSHLPTSSPAIESHIRIEGLSQKFIYLNDDVMFGKDWPPDF 420
QY 421 YSHSGQKQYLLTWVPVNCAGCGPSWIKDGYCDKACNNSACDWDGDCGNSGGSRYTAG 480
DB 421 YSHSGQKQYLLTWVPVNCAGCGPSWIKDGYCDKACNNSACDWDGDCGNSGGSRYTAG 480
QY 481 GGGTGSIGVGHPMQFGGINSVSYCNQGCANSMADKFCDOACNVLSGCFDAGDCQDHF 540
DB 481 GGGTGSIGVGHPMQFGGINSVSYCNQGCANSMADKFCDOACNVLSGCFDAGDCQDHF 540
QY 541 HELYKYLIPNQTHTIIPKGECLPYFSFAEVAKRGVAGYSDNPRIIRHASIANKMTIHL 600
DB 541 HELYKYLIPNQTHTIIPKGECLPYFSFAEVAKRGVAGYSDNPRIIRHASIANKMTIHL 600
QY 601 IMHSGNATTHFNLTFOQNTDEEFMOQITVEVDREGRKLNSTQKQYEMLVSTITLP 660
DB 601 IMHSGNATTHFNLTFOQNTDEEFMOQITVEVDREGRKLNSTQKQYEMLVSTITLP 660
QY 661 EAEIIPEDIPKKEKPEKPEKFDVNSTRAQOEVKIPLVNIISLPPDOALSTNTLDQIEH 720
DB 661 EAEIIPEDIPKKEKPEKPEKFDVNSTRAQOEVKIPLVNIISLPPDOALSTNTLDQIEH 720
QY 721 GQITLKGYLSSALRSFLMNSOHAQIKNOAILINDEFNDSLVAQOEQOVKSTILPNSLG 780
DB 721 GQITLKGYLSSALRSFLMNSOHAQIKNOAILINDEFNDSLVAQOEQOVKSTILPNSLG 780
QY 781 VSERLQRLTFPAVSVKNGHDQGNPDLLETTASRVEVTHQTKTIGGNVTKKPKPSLIV 840
DB 781 VSERLQRLTFPAVSVKNGHDQGNPDLLETTASRVEVTHQTKTIGGNVTKKPKPSLIV 840
```

```
DB 781 VSERLQRLTFPAVSVKNGHDQGNPDLLETTASRVEVTHQTKTIGGNVTKKPKPSLIV 840
QY 841 PLSQWTKKTKITGKEKENSMEENAHNIGVTEVLLGRDLQHYTDSYLGFLPWEKKKYF 900
DB 841 PLSQWTKKTKITGKEKENSMEENAHNIGVTEVLLGRDLQHYTDSYLGFLPWEKKKYF 900
QY 901 QDLIDEEESLKTQLAFTDSKNTGRQIK 928
DB 901 QDLIDEEESLKTQLAFTDSKNTGRQIK 928
```

```
RESULT 4
```

```
US-10-023-889-4
```

```
/ Sequence 4, Application US/10023889
```

```
/ Publication No. US20030124652A1
```

```
/ GENERAL INFORMATION:
```

```
/ APPLICANT: CANFIELD, William
```

```
/ TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARB
```

```
/ FILE REFERENCE: 203512US77
```

```
/ CURRENT APPLICATION NUMBER: US/10/023, 889
```

```
/ NUMBER OF SEQ ID NOS: 21
```

```
/ SOFTWARE: PatentIn version 3.1
```

```
/ SEQ ID NO 4
```

```
/ LENGTH: 928
```

```
/ TYPE: PRF
```

```
/ ORGANISM: Homo sapiens
```

```
US-10-023-889-4
```

```
Query Match 100.0%; Score 4907; DB 14; Length 928;
```

```
Best Local Similarity 100.0%; Pred. No. 0;
```

```
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MFKLORQTYTCLSHRYGLVYCFGLGVVYTVISAFQFGEVYLEWSDQYHVLFDSTYRDN 60
DB 1 MFKLORQTYTCLSHRYGLVYCFGLGVVYTVISAFQFGEVYLEWSDQYHVLFDSTYRDN 60
QY 61 AGKSFQNRCLPMPIDVYTVWNGTDLLELKELOQVREQMEBEQKAMREILGKNTTEPTK 120
DB 61 AGKSFQNRCLPMPIDVYTVWNGTDLLELKELOQVREQMEBEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLTHCIKYPMLVLDPALPANITLKDVPSLYPSFHSASDIPVAKKNSSTNV 180
DB 121 KSEKQLECLTHCIKYPMLVLDPALPANITLKDVPSLYPSFHSASDIPVAKKNSSTNV 180
QY 181 SYVVEPSTKDVDAHSGLLKGNRSQTVWNGYLTDDKEVGLVLMODLAFSGPPTFKET 240
DB 181 SYVVEPSTKDVDAHSGLLKGNRSQTVWNGYLTDDKEVGLVLMODLAFSGPPTFKET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFOELNKQTKKNTIDGKEITISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFOELNKQTKKNTIDGKEITISPA 300
QY 301 YLLMDLSAISQSKODEDISASRFEDNEBELRYSLSIERHAPVWRNIFIVTNGQIPSWNL 360
DB 301 YLLMDLSAISQSKODEDISASRFEDNEBELRYSLSIERHAPVWRNIFIVTNGQIPSWNL 360
QY 361 DNPRTVITVHQDVFNRLSHLPTSSPAIESHIRIEGLSQKFIYLNDDVMFGKDWPPDF 420
DB 361 DNPRTVITVHQDVFNRLSHLPTSSPAIESHIRIEGLSQKFIYLNDDVMFGKDWPPDF 420
QY 421 YSHSGQKQYLLTWVPVNCAGCGPSWIKDGYCDKACNNSACDWDGDCGNSGGSRYTAG 480
DB 421 YSHSGQKQYLLTWVPVNCAGCGPSWIKDGYCDKACNNSACDWDGDCGNSGGSRYTAG 480
QY 481 GGGTGSIGVGHPMQFGGINSVSYCNQGCANSMADKFCDOACNVLSGCFDAGDCQDHF 540
DB 481 GGGTGSIGVGHPMQFGGINSVSYCNQGCANSMADKFCDOACNVLSGCFDAGDCQDHF 540
QY 541 HELYKYLIPNQTHTIIPKGECLPYFSFAEVAKRGVAGYSDNPRIIRHASIANKMTIHL 600
DB 541 HELYKYLIPNQTHTIIPKGECLPYFSFAEVAKRGVAGYSDNPRIIRHASIANKMTIHL 600
```

Db 541 HELYKYLIPNQHYYIIPKGECLPYPSFAEVAKGVGAGYSDNPPIIRASIANKKTIHL 600
QY 601 IMHSGNATTIHNNLTFFONTNDEEFKMOITVEVDTRREGKLNSTAKGYENLVSPITLLP 660
Db 601 IMHSGNATTIHNNLTFFONTNDEEFKMOITVEVDTRREGKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPKEKRFKFRHDVNSTRAQOEVKIPLVNI SLPPDAQLSANTLDLQLEH 720
Db 661 EAEILFEDIPKEKRFKFRHDVNSTRAQOEVKIPLVNI SLPPDAQLSANTLDLQLEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSOHAKIKNOAIITDETNDLSVAPQEKQVHKSIIIPNSIG 780
Db 721 GDTTLKGYNLSSKALLRSFLMNSOHAKIKNOAIITDETNDLSVAPQEKQVHKSIIIPNSIG 780
QY 781 VSEIRLORLTFPAPVAVKVNHDGONPDLLETTAFRVEYTHQKTIIGVNTKEKPPSLIV 840
Db 781 VSEIRLORLTFPAPVAVKVNHDGONPDLLETTAFRVEYTHQKTIIGVNTKEKPPSLIV 840
QY 841 PLESQWTKKTKITGKEKENSMEENAEENHIGVTEVLGRKLOHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQWTKKTKITGKEKENSMEENAEENHIGVTEVLGRKLOHYTDSYLGFLPWEKKKYF 900
QY 901 QDLDEESLKTQLAFTDSKNTGRQK 928
Db 901 QDLDEESLKTQLAFTDSKNTGRQK 928
RESULT 5
US-10-023-890-4
; Sequence 4, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYD
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-890-4
Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFKLQROTTCLSHRVGLVYCFGVVVTIVSAFOGEBVLEMSRDQYHVLFDYSYRDI 60
Db 1 MFKLQROTTCLSHRVGLVYCFGVVVTIVSAFOGEBVLEMSRDQYHVLFDYSYRDI 60
QY 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELKELOQVREOMEBOQXAMREILGKNTTEPTK 120
Db 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELKELOQVREOMEBOQXAMREILGKNTTEPTK 120
QY 121 KSEKQLECLLTHCIRKPMVLVDPALPANITLKDVSILPSRHSASDINNVAKPKPSINV 180
Db 121 KSEKQLECLLTHCIRKPMVLVDPALPANITLKDVSILPSRHSASDINNVAKPKPSINV 180
QY 181 STVVEPSTKDVDAHSGLKNSRQTWMEGYLTDTKEVGLVIMODLAFLSGFPPTFKET 240
Db 181 STVVEPSTKDVDAHSGLKNSRQTWMEGYLTDTKEVGLVIMODLAFLSGFPPTFKET 240
QY 241 NOLKTKLPENLSSKYYKLLQYSEASVALIKLNPPDFQELNKQTKKNTITGKILITSPA 300
Db 241 NOLKTKLPENLSSKYYKLLQYSEASVALIKLNPPDFQELNKQTKKNTITGKILITSPA 300
QY 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRSHAPVVRNIFIVTNGQIPSMNL 360
Db 301 YLLMDLSAISQKODEDISASRFEDNEELRYSLSIRSHAPVVRNIFIVTNGQIPSMNL 360

QY 361 DNPRTVITHOVFRNLSHLPTSSPAIESHRIEGSOKFIYLNDDVMPFGKDVPPDF 420
Db 361 DNPRTVITHOVFRNLSHLPTSSPAIESHRIEGSOKFIYLNDDVMPFGKDVPPDF 420
QY 421 YSHSKQKXYLLTWPEVNCAGCGGSWIKDQCDKACNNASACMDGDCGSGSGRYIAG 480
Db 421 YSHSKQKXYLLTWPEVNCAGCGGSWIKDQCDKACNNASACMDGDCGSGSGRYIAG 480
QY 481 GGGTGSITGVHPQFQGGINSVYCNQGCANSLMDKFCDAQCNVLSGCFDAGDGGQDHF 540
Db 481 GGGTGSITGVHPQFQGGINSVYCNQGCANSLMDKFCDAQCNVLSGCFDAGDGGQDHF 540
QY 541 HELYKYLIPNQHYYIIPKGECLPYPSFAEVAKGVGAGYSDNPPIIRASIANKKTIHL 600
Db 541 HELYKYLIPNQHYYIIPKGECLPYPSFAEVAKGVGAGYSDNPPIIRASIANKKTIHL 600
QY 601 IMHSGNATTIHNNLTFFONTNDEEFKMOITVEVDTRREGKLNSTAKGYENLVSPITLLP 660
Db 601 IMHSGNATTIHNNLTFFONTNDEEFKMOITVEVDTRREGKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPKEKRFKFRHDVNSTRAQOEVKIPLVNI SLPPDAQLSANTLDLQLEH 720
Db 661 EAEILFEDIPKEKRFKFRHDVNSTRAQOEVKIPLVNI SLPPDAQLSANTLDLQLEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSOHAKIKNOAIITDETNDLSVAPQEKQVHKSIIIPNSIG 780
Db 721 GDTTLKGYNLSSKALLRSFLMNSOHAKIKNOAIITDETNDLSVAPQEKQVHKSIIIPNSIG 780
QY 781 VSEIRLORLTFPAPVAVKVNHDGONPDLLETTAFRVEYTHQKTIIGVNTKEKPPSLIV 840
Db 781 VSEIRLORLTFPAPVAVKVNHDGONPDLLETTAFRVEYTHQKTIIGVNTKEKPPSLIV 840
QY 841 PLESQWTKKTKITGKEKENSMEENAEENHIGVTEVLGRKLOHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQWTKKTKITGKEKENSMEENAEENHIGVTEVLGRKLOHYTDSYLGFLPWEKKKYF 900
QY 901 QDLDEESLKTQLAFTDSKNTGRQK 928
Db 901 QDLDEESLKTQLAFTDSKNTGRQK 928
RESULT 6
US-10-024-197-4
; Sequence 4, Application US/10024197
; Publication No. US2003013924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREOSIDASE AND METHODS
; FILE REFERENCE: 209794US0
; CURRENT APPLICATION NUMBER: US/10/024,197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-197-4
Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFKLQROTTCLSHRVGLVYCFGVVVTIVSAFOGEBVLEMSRDQYHVLFDYSYRDI 60
Db 1 MFKLQROTTCLSHRVGLVYCFGVVVTIVSAFOGEBVLEMSRDQYHVLFDYSYRDI 60
QY 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELKELOQVREOMEBOQXAMREILGKNTTEPTK 120
Db 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELKELOQVREOMEBOQXAMREILGKNTTEPTK 120

```
QY 121 KSEKQECCLTHGCIKYPMLVLDPALPANITLTKDVPSLVSFHSASDIFNVAKPKNPSTNV 180
DB 121 KSEKQECCLTHGCIKYPMLVLDPALPANITLTKDVPSLVSFHSASDIFNVAKPKNPSTNV 180
QY 181 SYVVPDSTKDVEDAHSGLLKGNRSQVTWGRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
DB 181 SYVVPDSTKDVEDAHSGLLKGNRSQVTWGRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
QY 241 NQKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFQELNKQTKKNTIDGKELTISPA 300
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFQELNKQTKKNTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLSRISERRAPWVRNIFIVTNGQIPSWML 360
DB 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLSRISERRAPWVRNIFIVTNGQIPSWML 360
QY 361 DNPRTVTVHODVFRNLSHLPTFSSPAIESHRIEGLSOKFIYLANDVMEGKDVPPDF 420
DB 361 DNPRTVTVHODVFRNLSHLPTFSSPAIESHRIEGLSOKFIYLANDVMEGKDVPPDF 420
QY 421 YSHSGQKYYLTPVPVNCAGCGGSGWIKDGYCDKACNNSACDWDGDCGSGSGSRYTAG 480
DB 421 YSHSGQKYYLTPVPVNCAGCGGSGWIKDGYCDKACNNSACDWDGDCGSGSGSRYTAG 480
QY 481 GGGTGSIGVGHFWQFGGINSVSYCNQGCANSLADKFCDAQCNVLSGFPDAGDGDHF 540
DB 481 GGGTGSIGVGHFWQFGGINSVSYCNQGCANSLADKFCDAQCNVLSGFPDAGDGDHF 540
QY 541 HELYKYLIPNQTHYIIPKGECLPYFSFAVAKRGVGAVSNDPIIRHASIANKKTIHL 600
DB 541 HELYKYLIPNQTHYIIPKGECLPYFSFAVAKRGVGAVSNDPIIRHASIANKKTIHL 600
QY 601 IMHSGMNAATTIHNLTFQNTNDEEFMQITVEVDTRREGPKLNSTQKGYENLVSPITLLP 660
DB 601 IMHSGMNAATTIHNLTFQNTNDEEFMQITVEVDTRREGPKLNSTQKGYENLVSPITLLP 660
QY 661 EABLLEFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSINTLDLQLEH 720
DB 661 EABLLEFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSINTLDLQLEH 720
QY 721 GDITLKGYNLSKALRSFLMNSQAKIKNOALITDETNDSLVAPQEKVHKSILPNSLG 780
DB 721 GDITLKGYNLSKALRSFLMNSQAKIKNOALITDETNDSLVAPQEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVNGHDGONPDLLETTARFRVETHQKTIIGNAVTEKEKPSLIIV 840
DB 781 VSERLQRLTFPAVSVKVNGHDGONPDLLETTARFRVETHQKTIIGNAVTEKEKPSLIIV 840
QY 841 PLESQMTKEKKITGKEKENSMEENAHNIGTVEVLGRKLOHYTDSYLGFLPWEKKKYF 900
DB 841 PLESQMTKEKKITGKEKENSMEENAHNIGTVEVLGRKLOHYTDSYLGFLPWEKKKYF 900
QY 901 QDLDEESLKTQLAFTDSKNTGRQLK 928
DB 901 QDLDEESLKTQLAFTDSKNTGRQLK 928
```

```
RESULT 7
US-10-023-894-4
; Sequence 4, Application US//10023894
; Publication No. US20030143699A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; APPLICANT: KORNFELD, Stuart
; TITLE OF INVENTION: EXPRESSION OF LYCOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; FILE REFERENCE: ACETYLGLUCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSIMANIDAS
; CURRENT APPLICATION NUMBER: US/10/023, 894
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 928
```

```
QY 121 KSEKQECCLTHGCIKYPMLVLDPALPANITLTKDVPSLVSFHSASDIFNVAKPKNPSTNV 180
DB 121 KSEKQECCLTHGCIKYPMLVLDPALPANITLTKDVPSLVSFHSASDIFNVAKPKNPSTNV 180
QY 181 SYVVPDSTKDVEDAHSGLLKGNRSQVTWGRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
DB 181 SYVVPDSTKDVEDAHSGLLKGNRSQVTWGRGYLTTRKEVPGVLMODLAFLSGFPPTFKET 240
QY 241 NQKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFQELNKQTKKNTIDGKELTISPA 300
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALKLNNPKDFQELNKQTKKNTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLSRISERRAPWVRNIFIVTNGQIPSWML 360
DB 301 YLLMDLSAISQSKODEDISASRFEDNEELRYSLSRISERRAPWVRNIFIVTNGQIPSWML 360
QY 361 DNPRTVTVHODVFRNLSHLPTFSSPAIESHRIEGLSOKFIYLANDVMEGKDVPPDF 420
DB 361 DNPRTVTVHODVFRNLSHLPTFSSPAIESHRIEGLSOKFIYLANDVMEGKDVPPDF 420
QY 421 YSHSGQKYYLTPVPVNCAGCGGSGWIKDGYCDKACNNSACDWDGDCGSGSGSRYTAG 480
DB 421 YSHSGQKYYLTPVPVNCAGCGGSGWIKDGYCDKACNNSACDWDGDCGSGSGSRYTAG 480
QY 481 GGGTGSIGVGHFWQFGGINSVSYCNQGCANSLADKFCDAQCNVLSGFPDAGDGDHF 540
DB 481 GGGTGSIGVGHFWQFGGINSVSYCNQGCANSLADKFCDAQCNVLSGFPDAGDGDHF 540
QY 541 HELYKYLIPNQTHYIIPKGECLPYFSFAVAKRGVGAVSNDPIIRHASIANKKTIHL 600
DB 541 HELYKYLIPNQTHYIIPKGECLPYFSFAVAKRGVGAVSNDPIIRHASIANKKTIHL 600
QY 601 IMHSGMNAATTIHNLTFQNTNDEEFMQITVEVDTRREGPKLNSTQKGYENLVSPITLLP 660
DB 601 IMHSGMNAATTIHNLTFQNTNDEEFMQITVEVDTRREGPKLNSTQKGYENLVSPITLLP 660
QY 661 EABLLEFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSINTLDLQLEH 720
DB 661 EABLLEFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSINTLDLQLEH 720
QY 721 GDITLKGYNLSKALRSFLMNSQAKIKNOALITDETNDSLVAPQEKVHKSILPNSLG 780
DB 721 GDITLKGYNLSKALRSFLMNSQAKIKNOALITDETNDSLVAPQEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVNGHDGONPDLLETTARFRVETHQKTIIGNAVTEKEKPSLIIV 840
DB 781 VSERLQRLTFPAVSVKVNGHDGONPDLLETTARFRVETHQKTIIGNAVTEKEKPSLIIV 840
QY 841 PLESQMTKEKKITGKEKENSMEENAHNIGTVEVLGRKLOHYTDSYLGFLPWEKKKYF 900
DB 841 PLESQMTKEKKITGKEKENSMEENAHNIGTVEVLGRKLOHYTDSYLGFLPWEKKKYF 900
QY 901 QDLDEESLKTQLAFTDSKNTGRQLK 928
DB 901 QDLDEESLKTQLAFTDSKNTGRQLK 928
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-894-4
```

```
Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT 8

```

US-10-306-686-1
; Sequence 1, Application US/10306686
; Publication No. US20030148460A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GlcNAcase OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 230397US77DIV
; CURRENT APPLICATION NUMBER: US/10/306,686
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: 09/636,596
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-08-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-306-686-1

```

Query Match 100.0%; Score 4907; DB 14; Length 928;

Best Local Similarity 100.0%; Pred. No. 0; Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MFPLIQRQTYTCLSHRGVLYCPIGVVVTIYSAFQFGBVLEMSRDQYHVLFDYSYRDI 60
DB 1 MFPLIQRQTYTCLSHRGVLYCPIGVVVTIYSAFQFGBVLEMSRDQYHVLFDYSYRDI 60
QY 61 AGKSFQNRCLPMPIDVYVYVWNGTDLLELKELOQVREOMEBOQAMSEIIQKNTTPTK 120
DB 61 AGKSFQNRCLPMPIDVYVYVWNGTDLLELKELOQVREOMEBOQAMSEIIQKNTTPTK 120
QY 121 KSEKOLECLTHCICIKVPMVLDPALPANITLKDVPSLYSPFASADIFNVAKPKXNSTNV 180
DB 121 KSEKOLECLTHCICIKVPMVLDPALPANITLKDVPSLYSPFASADIFNVAKPKXNSTNV 180
QY 121 KSEKOLECLTHCICIKVPMVLDPALPANITLKDVPSLYSPFASADIFNVAKPKXNSTNV 180
DB 121 KSEKOLECLTHCICIKVPMVLDPALPANITLKDVPSLYSPFASADIFNVAKPKXNSTNV 180
QY 181 SVVVPSTKDVDAHSGLLKNSRQTVWRGYLTTEKVPGLVMDLAFSLGPPPTFET 240
DB 181 SVVVPSTKDVDAHSGLLKNSRQTVWRGYLTTEKVPGLVMDLAFSLGPPPTFET 240
QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFOLNKQTKKNTIDGKELTISPA 300
DB 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFOLNKQTKKNTIDGKELTISPA 300
QY 301 YLLMDLSAISQKQEDDISASRFEDNEELRYSLSIRHAPWVRNIFIVTNGQIPSWNL 360
DB 301 YLLMDLSAISQKQEDDISASRFEDNEELRYSLSIRHAPWVRNIFIVTNGQIPSWNL 360
QY 361 DNPRTVIVHODVFNLSHLPTFSSPAIESHRIHRIEGLSOKFIYLNDDVMFGXVMPDF 420
DB 361 DNPRTVIVHODVFNLSHLPTFSSPAIESHRIHRIEGLSOKFIYLNDDVMFGXVMPDF 420
QY 421 YHSHKQKQVYLTWPVPCAGCGPGSWIKDGYCDKACNNACDMDGDCGSGSRIYAG 480
DB 421 YHSHKQKQVYLTWPVPCAGCGPGSWIKDGYCDKACNNACDMDGDCGSGSRIYAG 480
QY 481 GGTGSSIGVHPWQGGGINSYCNQGCANSMWLDKFCDOACNVLSGCFPAGDQGDHF 540
DB 481 GGTGSSIGVHPWQGGGINSYCNQGCANSMWLDKFCDOACNVLSGCFPAGDQGDHF 540
QY 541 HBLVYVILLPNQTHYIIPKGECLPYFSPFAVAKRGVEGAYSDNPIIRHSAIANRKTITL 600
DB 541 HBLVYVILLPNQTHYIIPKGECLPYFSPFAVAKRGVEGAYSDNPIIRHSAIANRKTITL 600
QY 601 INHSGNATTHIHNLTFOVNTDEEPMQIYEVDPREBPKXNSTOKXEYLVSPITLLP 660
DB 601 INHSGNATTHIHNLTFOVNTDEEPMQIYEVDPREBPKXNSTOKXEYLVSPITLLP 660
QY 661 EAEIIFEDIPEKEKPPKRGHDVNSTRAQEEVKIPLVNIISLPKDAQLSINTLDLQEH 720
DB 661 EAEIIFEDIPEKEKPPKRGHDVNSTRAQEEVKIPLVNIISLPKDAQLSINTLDLQEH 720

```

RESULT 9

```

US-10-023-888-2
; Sequence 2, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 203515US77
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: hybrid
US-10-023-888-2

```

Query Match 94.4%; Score 4630; DB 14; Length 1199;

Best Local Similarity 99.3%; Pred. No. 0; Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

QY 45 SRQCHVLPDSYRDNINAGSFPQNRCLPMPIDVYVYVWNGTDLLELKELOQVREOMEBO 104
DB 35 SRQCHVLPDSYRDNINAGSFPQNRCLPMPIDVYVYVWNGTDLLELKELOQVREOMEBO 94
QY 105 KAMREILGKNTTEPTKSEKOLECLTHCICIKVPMVLDPALPANITLKDVPSLYSPFSA 164
DB 95 KAMREILGKNTTEPTKSEKOLECLTHCICIKVPMVLDPALPANITLKDVPSLYSPFSA 154
QY 165 SDIFNVAKPKXNSTNVSVVVPSTKDVDAHSGLLKNSRQTVWRGYLTTEKVPGLVLM 224
DB 155 SDIFNVAKPKXNSTNVSVVVPSTKDVDAHSGLLKNSRQTVWRGYLTTEKVPGLVLM 214
QY 225 ODIAFLSGPPPTFETKTNOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFOLNKQ 284
DB 215 ODIAFLSGPPPTFETKTNOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFOLNKQ 274
QY 285 KKNMTIDGKELTISPAYLLMDLSAISQKQEDDISASRFEDNEELRYSLSIRHAPWVR 344
DB 275 KKNMTIDGKELTISPAYLLMDLSAISQKQEDDISASRFEDNEELRYSLSIRHAPWVR 334
QY 345 NIFIVNGQIPSWNLNDNPRVTVHODVFNLSHLPTFSSPAIESHRIHRIEGLSOKFIY 404
DB 335 NIFIVNGQIPSWNLNDNPRVTVHODVFNLSHLPTFSSPAIESHRIHRIEGLSOKFIY 394
QY 405 LNDVAFGMDVMPDDFVSHSKQKQVYLTWPVPCAGCGPGSWIKDGYCDKACNNACDMD 464
DB 395 LNDVAFGMDVMPDDFVSHSKQKQVYLTWPVPCAGCGPGSWIKDGYCDKACNNACDMD 454
QY 465 GGDGSGSGSRIYAGGGGTSIGVHPWQGGGINSYCNQGCANSMWLDKFCDOACN 524
DB 455 GGDGSGSGSRIYAGGGGTSIGVHPWQGGGINSYCNQGCANSMWLDKFCDOACN 514
QY 525 VLSGPDADPCGDHFHBLVYVILLPNQTHYIIPKGECLPYFSPFAVAKRGVEGAYSDNP 584

```


Db 515 VISCGRDADCCGDHFEHLYKVIILPNQTHYIIPGECIPYSEFAEVAKRGVAGASDNP 574
Qy 585 IIRHASIANKWKTIIHIMSGNATTTIHNLTFQNTNDEEFKQITVEVDTRREGPKLNT 644
Db 575 IIRHASIANKWKTIIHIMSGNATTTIHNLTFQNTNDEEFKQITVEVDTRREGPKLNT 634
Qy 645 AOKGYENLVSPITLLPEAEILFEDIPEKRFKFRKHDVNSTRAQEEVKIPLVNIISLLP 704
Db 635 AOKGYENLVSPITLLPEAEILFEDIPEKRFKFRKHDVNSTRAQEEVKIPLVNIISLLP 694
Qy 705 KDAQSLINTLDLQLEHGDITLKGYNLKSALLRSFLMNSQAKIKNOAIITDETNDSLVA 764
Db 695 KDAQSLINTLDLQLEHGDITLKGYNLKSALLRSFLMNSQAKIKNOAIITDETNDSLVA 754
Qy 765 POEKOVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPDLLETTARFVEVTHQK 824
Db 755 POEKOVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPDLLETTARFVEVTHQK 814
Qy 825 TIGGVNTEKRPSPILVPLESQMTKEKKTIGKEKENSMBENAMNHIQVTEVLGRKLOHY 884
Db 815 TIGGVNTEKRPSPILVPLESQMTKEKKTIGKEKENSMBENAMNHIQVTEVLGRKLOHY 874
Qy 885 TDSYLGFLPWEKKKTYFODLLDEESLKTQLAYFTDSKNRGR 925
Db 875 TDSYLGFLPWEKKKTYFODLLDEESLKTQLAYFTDSKNRGR 915

RESULT 10
US-10-023-889-2
; Sequence 2, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOH
; TITLE OF INVENTION: DEFICIENT CELLS
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PR
; ORGANISM: hybrid
US-10-023-889-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Qy 45 SRDQYHVLFDPSYNDINAGSFQNRCLPMPIDIVVYTWNGTDLELLKELOQVREQMEEQ 104
Db 35 SRDQYHVLFDPSYNDINAGSFQNRCLPMPIDIVVYTWNGTDLELLKELOQVREQMEEQ 94
Qy 105 KAMREILGNKTEPTKSEKOLECLLTHCICKVPMVLDPALPANITLKDVPISYSEFSA 164
Db 95 KAMREILGNKTEPTKSEKOLECLLTHCICKVPMVLDPALPANITLKDVPISYSEFSA 154
Qy 165 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQVWARGYLTDDKEVGLVLM 224
Db 155 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQVWARGYLTDDKEVGLVLM 214
Qy 225 QDLAFLSGPFPFKETNOJKTLPENLSKVKLLQLYSEASVALLKLNPKDFOELNKOT 284
Db 215 QDLAFLSGPFPFKETNOJKTLPENLSKVKLLQLYSEASVALLKLNPKDFOELNKOT 274
Qy 285 KKNMTIDGKELTISPAVLLMDLSAISQKQDDISASRPEDNEELRYLSRSTIERAAPVR 344
Db 275 KKNMTIDGKELTISPAVLLMDLSAISQKQDDISASRPEDNEELRYLSRSTIERAAPVR 334
Qy 345 NFIIVTNGQIPSWLNDNPRVTIIVTHQDVFRLNLSHLPFSSPAIESHRIEGLSQKFIY 404
Db 335 NFIIVTNGQIPSWLNDNPRVTIIVTHQDVFRLNLSHLPFSSPAIESHRIEGLSQKFIY 394

Db 335 NFIIVTNGQIPSWLNDNPRVTIIVTHQDVFRLNLSHLPFSSPAIESHRIEGLSQKFIY 394
Qy 405 LNDVWVGKDVMPDDFYSHSKGQKYYLTWVPNCAGECPSWTKDQYCDKACNNSACMD 464
Db 395 LNDVWVGKDVMPDDFYSHSKGQKYYLTWVPNCAGECPSWTKDQYCDKACNNSACMD 454
Qy 465 GSDCSGNSGGSRYIAGGGTGSIGVHPMOFGGINSVSVCNQGACNSWLADYFCQACN 524
Db 455 GSDCSGNSGGSRYIAGGGTGSIGVHPMOFGGINSVSVCNQGACNSWLADYFCQACN 514
Qy 525 VLSGPDAGCGGDHFEHLYKVIILPNQTHYIIPGECIPYSEFAEVAKRGVAGASDNP 584
Db 515 VLSGPDAGCGGDHFEHLYKVIILPNQTHYIIPGECIPYSEFAEVAKRGVAGASDNP 574
Qy 585 IIRHASIANKWKTIIHIMSGNATTTIHNLTFQNTNDEEFKQITVEVDTRREGPKLNT 644
Db 575 IIRHASIANKWKTIIHIMSGNATTTIHNLTFQNTNDEEFKQITVEVDTRREGPKLNT 634
Qy 645 AOKGYENLVSPITLLPEAEILFEDIPEKRFKFRKHDVNSTRAQEEVKIPLVNIISLLP 704
Db 635 AOKGYENLVSPITLLPEAEILFEDIPEKRFKFRKHDVNSTRAQEEVKIPLVNIISLLP 694
Qy 705 KDAQSLINTLDLQLEHGDITLKGYNLKSALLRSFLMNSQAKIKNOAIITDETNDSLVA 764
Db 695 KDAQSLINTLDLQLEHGDITLKGYNLKSALLRSFLMNSQAKIKNOAIITDETNDSLVA 754
Qy 765 POEKOVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPDLLETTARFVEVTHQK 824
Db 755 POEKOVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPDLLETTARFVEVTHQK 814
Qy 825 TIGGVNTEKRPSPILVPLESQMTKEKKTIGKEKENSMBENAMNHIQVTEVLGRKLOHY 884
Db 815 TIGGVNTEKRPSPILVPLESQMTKEKKTIGKEKENSMBENAMNHIQVTEVLGRKLOHY 874
Qy 885 TDSYLGFLPWEKKKTYFODLLDEESLKTQLAYFTDSKNRGR 925
Db 875 TDSYLGFLPWEKKKTYFODLLDEESLKTQLAYFTDSKNRGR 915

RESULT 11
US-10-023-890-2
; Sequence 2, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOH
; TITLE OF INVENTION: MAMMALIAN CELLS
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PR
; ORGANISM: hybrid
US-10-023-890-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Qy 45 SRDQYHVLFDPSYNDINAGSFQNRCLPMPIDIVVYTWNGTDLELLKELOQVREQMEEQ 104
Db 35 SRDQYHVLFDPSYNDINAGSFQNRCLPMPIDIVVYTWNGTDLELLKELOQVREQMEEQ 94
Qy 105 KAMREILGNKTEPTKSEKOLECLLTHCICKVPMVLDPALPANITLKDVPISYSEFSA 164
Db 95 KAMREILGNKTEPTKSEKOLECLLTHCICKVPMVLDPALPANITLKDVPISYSEFSA 154
Qy 165 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQVWARGYLTDDKEVGLVLM 224
Db 155 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQVWARGYLTDDKEVGLVLM 214

QY 225 QDLAFLSGPPTPEKTNQKTKL PENLSKVKLLQYSEASVALLKNNPKDFOELNKOT 284
DB 215 QDLAFLSGPPTPEKTNQKTKL PENLSKVKLLQYSEASVALLKNNPKDFOELNKOT 274
QY 285 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEBELRYSLSIRHAPVR 344
DB 275 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEBELRYSLSIRHAPVR 334
QY 345 NFIYVNGQIPSWMLNDNPRVTIVTHQVFNLSHLPTSSPAISHSIRIEGLSQKFIY 404
DB 335 NFIYVNGQIPSWMLNDNPRVTIVTHQVFNLSHLPTSSPAISHSIRIEGLSQKFIY 394
QY 405 LNDVWFGKDWPDDEFSHSGQKXYLLTPVPNCAGCGSWIKDGYCDKACNNSACMD 464
DB 395 LNDVWFGKDWPDDEFSHSGQKXYLLTPVPNCAGCGSWIKDGYCDKACNNSACMD 454
QY 465 GDDCSGSGSRYIAGGGTGSIGVGHPMQFGGINSVYCNQGCANSLADKFCDOACN 524
DB 455 GDDCSGSGSRYIAGGGTGSIGVGHPMQFGGINSVYCNQGCANSLADKFCDOACN 514
QY 525 VLSGCFDAGDCQGDHFEHLKYVILLPNQTHYIIPKGECLPYSPFAVAKGVEGAYSDNP 584
DB 515 VLSGCFDAGDCQGDHFEHLKYVILLPNQTHYIIPKGECLPYSPFAVAKGVEGAYSDNP 574
QY 585 IIRHASIANKKTIHILHMSGNMAATTIHNLTFQNTDNEEFKQITVEVDTRREGPKLNT 644
DB 575 IIRHASIANKKTIHILHMSGNMAATTIHNLTFQNTDNEEFKQITVEVDTRREGPKLNT 634
QY 645 AOKGYENLVSPITLLPEAEILFEDI PKERPFKFRHDVNSRRAOEYKILPVNISLLP 704
DB 635 AOKGYENLVSPITLLPEAEILFEDI PKERPFKFRHDVNSRRAOEYKILPVNISLLP 694
QY 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSPFLMSQHAKIKNOAIITDETNDLSVA 764
DB 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSPFLMSQHAKIKNOAIITDETNDLSVA 754
QY 765 POEKQVHKSIILPNSLGVSERLQRLTFPAVSVKVNGHDQONPDLLETTARFRVETHQX 824
DB 755 POEKQVHKSIILPNSLGVSERLQRLTFPAVSVKVNGHDQONPDLLETTARFRVETHQX 814
QY 825 TIGGANTKEKPPSLIVPLESOMTKKTKGKENSMBENNAENHIGTEVILGRLQHY 884
DB 815 TIGGANTKEKPPSLIVPLESOMTKKTKGKENSMBENNAENHIGTEVILGRLQHY 874
QY 885 TDSYLGFLPWEKKKXYFODLLDEEESLKTOLAYFTDSKNTGR 925
DB 875 TDSYLGFLPWEKKKXYFODLLDEEESLKTOLAYFTDSKNTGR 915

RESULT 12
US-10-024-197-2
; Sequence 2, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS C
; FILE REFERENCE: 209794USO
; CURRENT APPLICATION NUMBER: US/10/024,197
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PR
; ORGANISM: hybrid
US-10-024-197-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pseq. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQYHVLFDYSYNDIAGKSFQNRCLPMPIDVYTVWNGTDLBELIKELQVREOMEEO 104
DB 35 SRDQYHVLFDYSYNDIAGKSFQNRCLPMPIDVYTVWNGTDLBELIKELQVREOMEEO 94
QY 105 KAMEILGKNTTPTPKSEKOLECLTHCICVPMULDALPANITLKQVPSLYPSFHA 164
DB 95 KAMEILGKNTTPTPKSEKOLECLTHCICVPMULDALPANITLKQVPSLYPSFHA 154
QY 165 SDIFNVAKPKNPSTNSVVPFSTKQVDEAHSGLLKNGSRQVWGRYLTTRDXKVPGLVM 224
DB 155 SDIFNVAKPKNPSTNSVVPFSTKQVDEAHSGLLKNGSRQVWGRYLTTRDXKVPGLVM 214
QY 225 QDLAFLSGPPTPEKTNQKTKL PENLSKVKLLQYSEASVALLKNNPKDFOELNKOT 284
DB 215 QDLAFLSGPPTPEKTNQKTKL PENLSKVKLLQYSEASVALLKNNPKDFOELNKOT 274
QY 285 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEBELRYSLSIRHAPVR 344
DB 275 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRPEDNEBELRYSLSIRHAPVR 334
QY 345 NFIYVNGQIPSWMLNDNPRVTIVTHQVFNLSHLPTSSPAISHSIRIEGLSQKFIY 404
DB 335 NFIYVNGQIPSWMLNDNPRVTIVTHQVFNLSHLPTSSPAISHSIRIEGLSQKFIY 394
QY 405 LNDVWFGKDWPDDEFSHSGQKXYLLTPVPNCAGCGSWIKDGYCDKACNNSACMD 464
DB 395 LNDVWFGKDWPDDEFSHSGQKXYLLTPVPNCAGCGSWIKDGYCDKACNNSACMD 454
QY 465 GDDCSGSGSRYIAGGGTGSIGVGHPMQFGGINSVYCNQGCANSLADKFCDOACN 524
DB 455 GDDCSGSGSRYIAGGGTGSIGVGHPMQFGGINSVYCNQGCANSLADKFCDOACN 514
QY 525 VLSGCFDAGDCQGDHFEHLKYVILLPNQTHYIIPKGECLPYSPFAVAKGVEGAYSDNP 584
DB 515 VLSGCFDAGDCQGDHFEHLKYVILLPNQTHYIIPKGECLPYSPFAVAKGVEGAYSDNP 574
QY 585 IIRHASIANKKTIHILHMSGNMAATTIHNLTFQNTDNEEFKQITVEVDTRREGPKLNT 644
DB 575 IIRHASIANKKTIHILHMSGNMAATTIHNLTFQNTDNEEFKQITVEVDTRREGPKLNT 634
QY 645 AOKGYENLVSPITLLPEAEILFEDI PKERPFKFRHDVNSRRAOEYKILPVNISLLP 704
DB 635 AOKGYENLVSPITLLPEAEILFEDI PKERPFKFRHDVNSRRAOEYKILPVNISLLP 694
QY 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSPFLMSQHAKIKNOAIITDETNDLSVA 764
DB 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSPFLMSQHAKIKNOAIITDETNDLSVA 754
QY 765 POEKQVHKSIILPNSLGVSERLQRLTFPAVSVKVNGHDQONPDLLETTARFRVETHQX 824
DB 755 POEKQVHKSIILPNSLGVSERLQRLTFPAVSVKVNGHDQONPDLLETTARFRVETHQX 814
QY 825 TIGGANTKEKPPSLIVPLESOMTKKTKGKENSMBENNAENHIGTEVILGRLQHY 884
DB 815 TIGGANTKEKPPSLIVPLESOMTKKTKGKENSMBENNAENHIGTEVILGRLQHY 874
QY 885 TDSYLGFLPWEKKKXYFODLLDEEESLKTOLAYFTDSKNTGR 925
DB 875 TDSYLGFLPWEKKKXYFODLLDEEESLKTOLAYFTDSKNTGR 915

RESULT 13
US-10-023-894-2
; Sequence 2, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: hybrid
US-10-023-894-2
```

```
Query Match      94.4%; Score 4630; DB 14; Length 1199;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 45 SDDQYHVFDSYRDNIAGKSFONRLCLPMPIDVYTWNGTDLLELKELOQVREOMESEQ 104
DB 35 SDDQYHVFDSYRDNIAGKSFONRLCLPMPIDVYTWNGTDLLELKELOQVREOMESEQ 94
QY 105 KAMREILGKNTTEPTKSKSEKOLECLTHCIKVPMLVLPALPANITLKDVPGLVSPFSA 164
DB 95 KAMREILGKNTTEPTKSKSEKOLECLTHCIKVPMLVLPALPANITLKDVPGLVSPFSA 154
QY 165 SDIFVNAKPKNSTVSVVVPDSTKDVEDAHSGLLKGNRSQOTWVRGYLTDDKEVPGVLM 224
DB 155 SDIFVNAKPKNSTVSVVVPDSTKDVEDAHSGLLKGNRSQOTWVRGYLTDDKEVPGVLM 214
QY 225 QDLAFSGPPTFKETNOLKTLPENLISKVKYLQLYSEASVALLKANNPKDFOELNKOT 284
DB 215 QDLAFSGPPTFKETNOLKTLPENLISKVKYLQLYSEASVALLKANNPKDFOELNKOT 274
QY 285 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRHAPVVR 344
DB 275 KKNMTIDGKELTISPAYLLMDLSAISQSKODEDISASRFEDNEELRYSLSIRHAPVVR 334
QY 345 NIFVTNGQIPSNLMDNPRVTIVTHQVFNLSHPFTSSPAISHSIRIEGJSQKTY 404
DB 335 NIFVTNGQIPSNLMDNPRVTIVTHQVFNLSHPFTSSPAISHSIRIEGJSQKTY 394
QY 405 LNDVDFGKDVDPDDPYSHSGOKYLLTFVPVNCAGEGSGSIKDGCKAKNNSACMD 464
DB 395 LNDVDFGKDVDPDDPYSHSGOKYLLTFVPVNCAGEGSGSIKDGCKAKNNSACMD 454
QY 465 GDDCGSGSGSRYIAGGGGTGSIIGVGHPMQFGGINSVSYCQGCANSWMLADKFCDAQN 524
DB 455 GDDCGSGSGSRYIAGGGGTGSIIGVGHPMQFGGINSVSYCQGCANSWMLADKFCDAQN 514
QY 525 VTSQGFADGCGODHHELYKVTLLPNOCHYIIPKGECLPYSPFAVAKRGVEGAYSDNP 584
DB 515 VTSQGFADGCGODHHELYKVTLLPNOCHYIIPKGECLPYSPFAVAKRGVEGAYSDNP 574
QY 585 IIRHNSIANKMTIHLIMHSGNNAATTIHENLFTONTDEEFMQITVEVDTREBGLNST 644
DB 575 IIRHNSIANKMTIHLIMHSGNNAATTIHENLFTONTDEEFMQITVEVDTREBGLNST 634
QY 645 AOKGYNVNLSPTILLPEAEILFEDIPKPKFRKHDVNSTRAQOEYKIPLVNISLLP 704
DB 635 AOKGYNVNLSPTILLPEAEILFEDIPKPKFRKHDVNSTRAQOEYKIPLVNISLLP 694
QY 705 KAOGLSLNTLQLEHGDITLKGYNLSKGLRSFAMNSQAKINQOAIITDETNDSLVA 764
DB 695 KAOGLSLNTLQLEHGDITLKGYNLSKGLRSFAMNSQAKINQOAIITDETNDSLVA 754
QY 765 POBKQVHKSILPNSLGVSERLQRLTFPAVSXVNGHGOQNPDLLETTARFRVETHTOK 824
DB 755 POBKQVHKSILPNSLGVSERLQRLTFPAVSXVNGHGOQNPDLLETTARFRVETHTOK 814
QY 825 TIGGANTKPKPSLIAVPLESOMTKEKKITGKEKENSMBENNAENHIGTEVLLGKRLQHY 884
DB 815 TIGGANTKPKPSLIAVPLESOMTKEKKITGKEKENSMBENNAENHIGTEVLLGKRLQHY 874
QY 885 TDSYGLFLPWEKKKYFODLLDEEESIKTOLAVFTDSKNTGR 925
DB 875 TDSYGLFLPWEKKKYFODLLDEEESIKTOLAVFTDSKNTGR 915
```

```
RESULT 14
US-10-120-801-88
; Sequence 88; Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Kerkuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Ruad
; APPLICANT: Topper, James N.
; APPLICANT: Malysankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smitheon, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLES OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 88
; LENGTH: 847
; TYPE: PRT
; ORGANISM: human
US-10-120-801-88

Query Match      91.2%; Score 4473; DB 15; Length 847;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 845; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MLFELLORQTYTCLSHRYGLVCFGLGVVTVISAFQGEVLEWSDQYHVFDSYRDNI 60
DB 1 MLFELLORQTYTCLSHRYGLVCFGLGVVTVISAFQGEVLEWSDQYHVFDSYRDNI 60
QY 61 AGKSFONRLCLPMPIDVYTWNGTDLLELKELOQVREOMESEQKAMREILGKNTTEPTK 120
DB 61 AGKSFONRLCLPMPIDVYTWNGTDLLELKELOQVREOMESEQKAMREILGKNTTEPTK 120
QY 121 KSEKOLECLTHCIKVPMLVLPALPANITLKDVPGLVSPFHSASDIENVAKPKNSTV 180
DB 121 KSEKOLECLTHCIKVPMLVLPALPANITLKDVPGLVSPFHSASDIENVAKPKNSTV 180
QY 181 SVVVFDSKDVEDAHSGLLKGNRSQOTWVRGYLTDDKEVPLVMODLAFSGPPTFKET 240
DB 181 SVVVFDSKDVEDAHSGLLKGNRSQOTWVRGYLTDDKEVPLVMODLAFSGPPTFKET 240
QY 241 NOLKTKLPENLSKVYLLQLYSEASVALLKANNPKDFOELNKOTKKNMTIDGKELTISPA 300
DB 241 NOLKTKLPENLSKVYLLQLYSEASVALLKANNPKDFOELNKOTKKNMTIDGKELTISPA 300
QY 301 YLMDLSAISQSKODEDISASRFEDNEELRYSLSIRHAPVVRNIFVTNGQIPSWMLN 360
DB 301 YLMDLSAISQSKODEDISASRFEDNEELRYSLSIRHAPVVRNIFVTNGQIPSWMLN 360
```

QY 361 DNRVITVTHQVFRNLSHLPTFSSPAISHHRIEGLSOKRTTYLNDVMPGKDWPDDE 420
DB 361 DNRVITVTHQVFRNLSHLPTFSSPAISHHRIEGLSOKRTTYLNDVMPGKDWPDDE 420
QY 421 YSHSKQKXYLLTWPVNCABGCGPSWIKDGYCDKACNNACDMDGDCSGNSGSRYYAG 480
DB 421 YSHSKQKXYLLTWPVNCABGCGPSWIKDGYCDKACNNACDMDGDCSGNSGSRYYAG 480
QY 481 GGGTSGIGVGHWPQFGGINSVSYCNQGCANSMWLDKFCDOACNVLSCGFDAGCGDHF 540
DB 481 GGGTSGIGVGHWPQFGGINSVSYCNQGCANSMWLDKFCDOACNVLSCGFDAGCGDHF 540
QY 541 HELYKYLIPNQTHTYIIPKGECLPYPSFAEVAKRGEAGYSDNPRIIRASIANKKTHL 600
DB 541 HELYKYLIPNQTHTYIIPKGECLPYPSFAEVAKRGEAGYSDNPRIIRASIANKKTHL 600
QY 601 IMHSGNNAATTIHFNLTFOQNTNDEEFKMOITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660
DB 601 IMHSGNNAATTIHFNLTFOQNTNDEEFKMOITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPEKRRPFKFRHDVNSTRAQEEVKIPLVNTISLLPKDAQSLNTLDLQEH 720
DB 661 EAEILFEDIPEKRRPFKFRHDVNSTRAQEEVKIPLVNTISLLPKDAQSLNTLDLQEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIITDETNDSLVAPQEKVHKSILPNSLG 780
DB 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIITDETNDSLVAPQEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVNHDGQNPDLDETTAFRVEYTHQKTIIGAVTKEKPSLIY 840
DB 781 VSERLQRLTFPAVSVKVNHDGQNPDLDETTAFRVEYTHQKTIIGAVTKEKPSLIY 840
QY 841 PLESQWT 847
DB 841 PLESQWT 847

RESULT 15
US-10-094-749-2392
; Sequence 2392, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YUKI
; APPLICANT: OTSUKA, KAGURU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL, FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2392
; LENGTH: 846
; TYPE: PRT

; ORGANISM: Homo sapiens
US-10-094-749-2392
Query Match 91.1%; Score 4468; DB 15; Length 846;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 844; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLEKLLQRTYTCLSHRGGLVYCFLLAVVYTYISAFQGVVLWMSRDQYHVLFDSDYRNI 60
DB 1 MLEKLLQRTYTCLSHRGGLVYCFLLAVVYTYISAFQGVVLWMSRDQYHVLFDSDYRNI 60
QY 61 AGSFQNRCLPMPIDVYVTVNGTDLLELLKELOQVREOMBEOKMREILGNTEPRK 120
DB 61 AGSFQNRCLPMPIDVYVTVNGTDLLELLKELOQVREOMBEOKMREILGNTEPRK 120
QY 121 KSEKQLECLTHCICIKPMVLDPALPANTLKDLPSELVPSFHSASDIENVAKPKNSTNV 180
DB 121 KSEKQLECLTHCICIKPMVLDPALPANTLKDLPSELVPSFHSASDIENVAKPKNSTNV 180
QY 181 SVVVPSTQDVEPAHSGLLKGNRQRTWMEGYLTTDKEVGVLMODLAFSGPPPFKEET 240
DB 181 SVVVPSTQDVEPAHSGLLKGNRQRTWMEGYLTTDKEVGVLMODLAFSGPPPFKEET 240
QY 241 NOLKTKLPENLSSKVLLOLYSEASVALLKNNPKDFQELNQTKKMTIDGKELTISPA 300
DB 241 NOLKTKLPENLSSKVLLOLYSEASVALLKNNPKDFQELNQTKKMTIDGKELTISPA 300
QY 301 YLLMDLSAISQSKODDIIASRPEDEELRYSLSRERAPVWRNIFTYNGQISMLNL 360
DB 301 YLLMDLSAISQSKODDIIASRPEDEELRYSLSRERAPVWRNIFTYNGQISMLNL 360
QY 361 DNRVITVTHQVFRNLSHLPTFSSPAISHHRIEGLSOKRTTYLNDVMPGKDWPDDE 420
DB 361 DNRVITVTHQVFRNLSHLPTFSSPAISHHRIEGLSOKRTTYLNDVMPGKDWPDDE 420
QY 421 YSHSKQKXYLLTWPVNCABGCGPSWIKDGYCDKACNNACDMDGDCSGNSGSRYYAG 480
DB 421 YSHSKQKXYLLTWPVNCABGCGPSWIKDGYCDKACNNACDMDGDCSGNSGSRYYAG 480
QY 481 GGGTSGIGVGHWPQFGGINSVSYCNQGCANSMWLDKFCDOACNVLSCGFDAGCGDHF 540
DB 481 GGGTSGIGVGHWPQFGGINSVSYCNQGCANSMWLDKFCDOACNVLSCGFDAGCGDHF 540
QY 541 HELYKYLIPNQTHTYIIPKGECLPYPSFAEVAKRGEAGYSDNPRIIRASIANKKTHL 600
DB 541 HELYKYLIPNQTHTYIIPKGECLPYPSFAEVAKRGEAGYSDNPRIIRASIANKKTHL 600
QY 601 IMHSGNNAATTIHFNLTFOQNTNDEEFKMOITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660
DB 601 IMHSGNNAATTIHFNLTFOQNTNDEEFKMOITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660
QY 661 EAEILFEDIPEKRRPFKFRHDVNSTRAQEEVKIPLVNTISLLPKDAQSLNTLDLQEH 720
DB 661 EAEILFEDIPEKRRPFKFRHDVNSTRAQEEVKIPLVNTISLLPKDAQSLNTLDLQEH 720
QY 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIITDETNDSLVAPQEKVHKSILPNSLG 780
DB 721 GDTTLKGYNLSSKALLRSFLMNSQAKIKNOAIITDETNDSLVAPQEKVHKSILPNSLG 780
QY 781 VSERLQRLTFPAVSVKVNHDGQNPDLDETTAFRVEYTHQKTIIGAVTKEKPSLIY 840
DB 781 VSERLQRLTFPAVSVKVNHDGQNPDLDETTAFRVEYTHQKTIIGAVTKEKPSLIY 840
QY 841 PLESQWT 846
DB 841 PLESQWT 846

Search completed: July 26, 2004, 11:25:25
Job time : 74.3389 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:58 ; Search time 15.4568 Seconds
(without alignments)
3126.212 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907
Sequence: 1 MFLKLRQRTYTCISHRYGL.....SLKTQLAVFTDSKNTGRQLK 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168.5	3.4	2470	1 NTC2_MOUSE	Q35516 mus musculus
2	168.5	3.4	2471	1 NTC2_HUMAN	Q04721 homo sapien
3	168.5	3.4	2471	1 NTC2_RAT	Q96W30 rattus norv
4	166.5	3.4	1429	1 L112_CAEL	P14585 caenorhabdi
5	165	3.4	2531	1 NTC1_RAT	Q07008 rattus norv
6	163	3.3	2531	1 NTC1_MOUSE	Q01705 mus musculu
7	162.5	3.3	2437	1 NTC1_BRARE	P46530 brachydantio
8	154.5	3.1	2321	1 NTC3_HUMAN	Q9um47 homo sapien
9	152	3.1	1664	1 NTC4_MOUSE	P31695 mus musculu
10	151.5	3.1	2003	1 NTC4_HUMAN	Q99466 homo sapien
11	149.5	3.0	2224	1 NTC3_XENLA	P21783 xenopus lae
12	148	3.0	2319	1 NTC3_RAT	Q9i172 rattus norv
13	147	3.0	2318	1 NTC3_MOUSE	Q61982 mus musculu
14	145.5	3.0	2556	1 NTC1_HUMAN	P46531 homo sapien
15	145	3.0	626	1 GIDA_BUCBP	P59485 buchnera ap
16	137.5	2.8	1328	1 YMO1_YEAST	Q04670 saccharomyc
17	136.5	2.8	1701	1 MSP1_PLAIF	P08569 plasmodium
18	135.5	2.8	1701	1 MSP1_PLAIF	P18819 plasmodium
19	134.5	2.7	2434	1 YCP1_OENHO	Q9mth5 oenothera h
20	134	2.7	1295	1 GLP1_CAEEL	P13508 caenorhabdi
21	131	2.7	2095	1 RPI_MOUSE	P56116 mus musculu
22	129.5	2.6	1630	1 MSP1_PLAIF	P04932 plasmodium
23	129.5	2.6	1630	1 MSP1_PLAIF	P04933 plasmodium
24	129.5	2.6	2867	1 RBP2_PLAIV	Q00799 plasmodium
25	129	2.6	2703	1 NTC2_DROME	P07207 drosophila
26	128.5	2.6	756	1 Y328_NYCGE	Q94199 mycoplasma
27	128	2.6	631	1 GIDA_BUCBP	O51879 buchnera ap
28	127.5	2.6	780	1 MUS2_BORBU	O51125 buchejia bu
29	127	2.6	1459	1 GEAT_YEAST	P39993 saccharomyc
30	126.5	2.6	1002	1 CLNN_HUMAN	Q961q2 homo sapien
31	126.5	2.6	1328	1 YMT5_YEAST	Q04214 saccharomyc
32	125.5	2.6	2738	1 PGCV_RAT	Q9ebh4 rattus norv
33	124	2.5	3135	1 S230_PLAFO	Q08372 plasmodium

ALIGNMENTS

RESULT 1	NTC2_MOUSE	STANDARD;	PRT; 2470 AA.
AC	O35516; Q06008; Q60941;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch		
DE	B).		
GN	NOTCH2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=Thymus;		
RA	Hamada Y., Higuchi M., Tsujimoto Y.;		
RT	"Complete amino acid sequence and multicistronic transcripts encoded by a		
RT	single copy of mouse Notch2 gene."		
RL	Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE OF 316-1518 FROM N.A.		
RC	STRAIN=C57BL/6 X CBA; TISSUE=Embryo;		
RX	MEDLINE=93178563; PubMed=8440332;		
RA	Lardelli M., Lendahl U.;		
RT	"Notch A and Notch B-two mouse Notch homologues coexpressed in a		
RT	wide variety of tissues."		
RL	Exp. Cell Res. 204:364-372(1993).		
RN	[3]		
RP	SEQUENCE OF 1765-2153 FROM N.A.		
RX	MEDLINE=97075110; PubMed=8917536;		
RA	Malner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,		
RT	Martin D.I.;		
RT	"Inhibition of granulocytic differentiation by mNotch1.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).		
RN	[4]		
RP	FUNCTION		
RX	MEDLINE=9936706; PubMed=10393120;		
RA	Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,		
RT	Tsujimoto Y.;		
RT	"Mutation in ankyrin repeats of the mouse Notch2 gene induces early		
RT	embryonic lethality."		
RL	Development 126:3415-3424(1999).		
RN	[5]		
RP	DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.		
RX	MEDLINE=95333893; PubMed=7609614;		
RA	Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;		
RT	"differential expression of Notch1 and Notch2 in developing and adult		
RT	mouse brain."		
RL	Brain Res. Mol. Brain Res. 23:263-272(1995).		
RN	[6]		
RP	POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.		
RX	MEDLINE=21523956; PubMed=11518718;		
RA	Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;		
RT	"Marine notch homologs (N1-4) undergo presenilin-dependent		

34	124	2.5	3259	1	G031_HUMAN	Q14789 homo sapien
35	123	2.5	770	1	DBP4_YEAST	P20448 saccharomyc
36	123	2.5	5430	1	MACE_HUMAN	Q9ubn3 homo sapien
37	122.5	2.5	1111	1	EXSC_BUCBP	Q96ab4 buchnera ap
38	122.5	2.5	1465	1	SPAZ_YEAST	P23201 saccharomyc
39	122.5	2.5	1541	1	YHVE_YEAST	P38873 saccharomyc
40	122.5	2.5	2763	1	TEBG_VZVD	P09278 varicella-z
41	122	2.5	1312	1	PA50_YEAST	P12753 saccharomyc
42	122	2.5	5938	1	MACE_HUMAN	Q96pk2 homo sapien
43	121.5	2.5	1755	1	YJ25_YEAST	P47100 saccharomyc
44	121.5	2.5	2025	1	TTG3_HUMAN	P53804 homo sapien
45	121	2.5	1052	1	CLNN_MOUSE	Q8c5w0 mus musculu

FT	DISUPLID	248	257	BY SIMILARITY.
Query Match		3.4%	Score 168.5; DB 1; Length 2471;	
Best Local Similarity		19.7%	Pred. No. 0.042;	
Matches	114;	Conservative	66; Mismatches	165; Indels 215; Gaps 27;
QY	434	EVPCNACBSCPSGSMWIDGYCDYACNNSACDWDGDCSGNGSGSRYIAGCGGTGSGVGHFW	493	
DB	1421	PPATCTISQYCAKDRKARDGVCBACNSHACQWGDGDC	1463	
QY	494	-QFGGAGSVASYCNOGC-----ANS-----WLDKF-----CDQACN	524	
DB	1464	ANCSPLPCMYINNOGDELCTNYECLFDNECCGNSCTCKYDKACDHFNDHNGCNC	1522	
QY	525	VLSGCPADGCGDGFHELYK-----VLLPNQTHYII PKBCLPYSPFAEV-----	571	
DB	1584	SEEGCGMGLDCAADQEPENIAEGTLVIVLMF-----PEQLLODARSFLRALTILHTN	1576	
QY	572	--ARKGVGAASDPIIRHSAIAKMKTIHLIMHSGNATTIHNLPQNTNDEEF---	626	
DB	1577	LRIKRDGSGELMVPYFYGKSAANKKQ-----RMTRRSLPGEQOEVA	1619	
QY	627	-MQITVEDTREGFKLNSIAKQGVENLSPITLPEABILFEDIPKEXRPFKFRHDVNS	685	
DB	1620	GSKFVLEIDNRQ-----CYQDSHCCKNTDAAAILAS-----HAIQG	1657	
QY	686	TRRQAEVKIPLVNI-----SLPKAQLSLNTLDQL-----EKDIT	724	
DB	1658	T-----LSPLVLVSVESESLPEPTOL-LYLLAVAVIIFILLGVIMAKRKHOSLW	1710	
QY	725	L-KGVNLSKSLALSPFMSNSHAK-----IKQALITDETNSLVAPOEKVHKS	773	
DB	1711	LPBEFTLRDRD-----SNHRRREPVQDAGVGLKSLVSQBSAN-LI-----	1750	
QY	774	ILPNSLGVSERLQSLTPPAVSVKVNGHDG---QNPPLDETTARFRVETHIOKTIQGN	829	
DB	1751	---GTGSEHWVDDESGPOPK-KYKABDEALISEEDPIDR-----BPWTQOHBA	1797	
QY	830	VTKKZPEPSLIYPLPESQWTKKTKTGKEXENSRMEKNENHIGVTEVLLGRQLQYTDYSL	889	
DB	1798	DIREFTSLALTPPA-----EQEVDVLIVNVRGPDGCTPMLA-----	1835	
QY	890	GFLPWEKKRYFQDILLDEESLKTQLA-YFTDSKNTGRQK	928	
DB	1836	-----SLRGSSDLSDDEDADAEDSSANITIDLVYQGSALQ	1870	
RESULT 3				
NTC2_RAT				
ID	NTC2_RAT	STANDARD;	PRT; 2471 AA.	
AC	O9QW30;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Neurogenic locus notch homolog protein 2 precursor (Notch 2).			
GN	NOTCH2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
NCBI	TaxID=10116;			
QX	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain.			
RC	TISSUE=Brain.			
RX	MEDLINE=93202015; PubMed=1295745;			
RT	Weinmaster G., Roberts V.J., Lemke G.;			
RT	"Notch2: a second mammalian Notch gene.";			
RL	Development 116:931-941 (1992).			
RT	[2]			
RP	TISSUE SPECIFICITY.			
RC	MEDLINE=21331789; PubMed=11438922;			
RA	Iryin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;			
RT	"Expression patterns of Notch1, Notch2, and Notch3 suggest multiple			
RT	functional roles for the Notch-DSL signaling system during brain			

RT development.":
RL J. Comp. Neurol. 436:167-181(2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs. May play an essential role in postimplantation
CC development, probably in some aspect of cell specification and/or
CC differentiation (By similarity).
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid
CC plexus in the brain. Expressed in postnatal central nervous system
CC (CNS) germinal zones and, in early postnatal life, within numerous
CC cells throughout the CNS. It is more highly localized to
CC ventricular germinal zones. Also found in the heart, liver and
CC kidney.
CC -1- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 35 EGF-like domains.
CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SMS-S-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M93661; AK13558.1; -.
DR PIR: A49128; A49128.
DR HSSP: P00743; ICGF.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx hydroxyl_S.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_11.
DR InterPro: IPR006209; EGF_11ke.
DR InterPro: IPR002049; LamInin_EGF.
DR InterPro: IPR008297; Notch.
DR InterPro: IPR000800; Notch_dom.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 35.
DR Pfam: PF00066; notch; 2.
DR PIRSF: PIRSF02279; Notch; 1.
DR PRINTS: PR00010; EGFBL00D.
DR PRINTS: PR00011; EGFAMININ.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 6.
DR SMART: SM00179; EGF_CA; 24.
DR SMART: SM00004; NU; 2.
DR PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PSS0088; ANK_REPEAT; 4.
DR PROSITE: PSS0010; ASX_HYDROXYL; 22.

DR PROSITE; PSS0022; EGF_1; 34.
DR PROSITE; PSS0186; EGF_2; 26.
DR PROSITE; PSS0026; EGF_3; 35.
DR PROSITE; PSS0187; EGF_CA; 22.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 142
FT DISULFID 148 159

FT SIGNAL 1 25
FT CHAIN 26 2471
FT CHAIN 1666 2471
FT CHAIN 1697 2471
FT DOMAIN 26 1677
FT TRANSSEM 1678 1698
FT DOMAIN 1699 2471
FT DOMAIN 26 63
FT DOMAIN 64 102
FT DOMAIN 105 143
FT DOMAIN 144 180
FT DOMAIN 182 219
FT DOMAIN 221 238
FT DOMAIN 260 296
FT DOMAIN 298 336
FT DOMAIN 338 374
FT DOMAIN 375 413
FT DOMAIN 415 454
FT DOMAIN 456 492
FT DOMAIN 494 530
FT DOMAIN 532 568
FT DOMAIN 570 605
FT DOMAIN 607 643
FT DOMAIN 645 680
FT DOMAIN 682 718
FT DOMAIN 720 755
FT DOMAIN 757 793
FT DOMAIN 795 831
FT DOMAIN 833 871
FT DOMAIN 873 909
FT DOMAIN 911 947
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1061
FT DOMAIN 1063 1099
FT DOMAIN 1101 1147
FT DOMAIN 1149 1185
FT DOMAIN 1187 1223
FT DOMAIN 1225 1262
FT DOMAIN 1264 1302
FT DOMAIN 1304 1343
FT DOMAIN 1374 1412
FT DOMAIN 1445 1482
FT DOMAIN 1494 1535
FT DOMAIN 1594 1635
FT DOMAIN 1645 1682
FT DOMAIN 1686 1723
FT REPEAT 1827 1871
FT REPEAT 1876 1905
FT REPEAT 1909 1939
FT REPEAT 1943 1972
FT REPEAT 1976 2005
FT REPEAT 2009 2038
FT DISULFID 28 41
FT DISULFID 35 51
FT DISULFID 53 62
FT DISULFID 68 79
FT DISULFID 73 90
FT DISULFID 92 101
FT DISULFID 109 121
FT DISULFID 115 131
FT DISULFID 133 14

FT DISULFID 153 168 BY SIMILARITY.
 FT DISULFID 170 179 BY SIMILARITY.
 FT DISULFID 196 198 BY SIMILARITY.
 FT DISULFID 192 207 BY SIMILARITY.
 FT DISULFID 209 218 BY SIMILARITY.
 FT DISULFID 225 236 BY SIMILARITY.
 FT DISULFID 230 246 BY SIMILARITY.
 FT DISULFID 248 257 BY SIMILARITY.
 FT DISULFID 264 275 BY SIMILARITY.
 FT DISULFID 269 284 BY SIMILARITY.
 FT DISULFID 286 295 BY SIMILARITY.
 FT DISULFID 302 315 BY SIMILARITY.
 FT DISULFID 309 324 BY SIMILARITY.
 FT DISULFID 326 335 BY SIMILARITY.
 FT DISULFID 342 353 BY SIMILARITY.
 FT DISULFID 347 362 BY SIMILARITY.
 FT DISULFID 364 373 BY SIMILARITY.
 FT DISULFID 379 380 BY SIMILARITY.
 FT DISULFID 384 401 BY SIMILARITY.
 FT DISULFID 403 412 BY SIMILARITY.
 FT DISULFID 419 433 BY SIMILARITY.
 FT DISULFID 427 442 BY SIMILARITY.
 FT DISULFID 444 453 BY SIMILARITY.
 FT DISULFID 460 471 BY SIMILARITY.
 FT DISULFID 465 480 BY SIMILARITY.
 FT DISULFID 482 481 BY SIMILARITY.
 FT DISULFID 498 509 BY SIMILARITY.
 FT DISULFID 503 518 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 536 547 BY SIMILARITY.

Query Match 3.4%; Score 168.5; DB 1; Length 2471;
 Best Local Similarity 26.2%; Pred. No. 0.042; 58; Indels 51; Gaps 5;

Matches 44; Conservative 15; Mismatches 5;
 QY 417 PDDFYSHSKQKRYVITWVPCACBQCPGSGWINDGYCDKACNNACDMWDGCGSGNSGSR 476
 DB 1404 PEFWGHSGSYAPRSTPRATCLSGYCADKXADGICDEACNASHAQWMDGDC----- 1455
 QY 477 YIAGGGGTSIGVGPW-QFGGINSVSVCNGC-----AN 511
 DB 1456 -----SLTMEBPMANCTSSILNCWEYINNQCDELCTAECLEDFNECGRNSKTCKYD 1506
 QY 512 SWIADKFE-----CDQACNVLSGCFDAGCGODHFEHLYK-----VILLP 550
 DB 1507 KYCAHDFKDHCKKCNNECGWDLDCADQPEVLAIGILVTVLLP 1554

RESULT 4
 L112 CAEBL STANDARD; PRT; 1429 AA.
 AC P14585;
 DT 01-JAN-1990 (rel. 13, Created)
 DT 01-JAN-1990 (rel. 13, Last sequence update)
 DT 15-MAR-2004 (rel. 43, Last annotation update)
 DE Lin-12 protein precursor.
 GN Lin-12 OR R107.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OK NCBI_Taxid=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Brictol N2;
 RX MEDLINE=88334747; PubMed=3419531;
 RA Yochem J., Weston K., Greenwald I.,
 RT "The Caenorhabditis elegans lin-12 gene encodes a transmembrane
 RL protein with overall similarity to Drosophila Notch.",
 RL Nature 335:547-550(1988).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Brictol N2;
 RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainecough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kristen J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thery-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL elegans.";
 RL Nature 368:32-38(1994).
 CC -1- FUNCTION: Involved in several cell fates decisions that requires
 CC cell-cell interactions. It is possible that lin-12 encodes a
 CC membrane-bound receptor for a signal that enables expression of
 CC the ventral uterine precursor cell fate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.
 CC -1- SIMILARITY: Contains 13 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M12069; AAA70191.1; -;
 DR EMBL: Z14092; CAAY8474.1; -;
 DR PIR: S06434; S06434.
 DR HSP: P00740; 1EDM.
 DR WormRep: R107.8; CE00274.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx hydroxyl_S.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001861; EGF_Ca.
 DR InterPro: IPR006209; EGF-like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR000800; Notch_dom.
 DR Pfam: PF00023; ank; 6.
 DR Pfam: PF00008; EGF; 13.
 DR Pfam: PF00066; notch; 3.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PRINTS: PR01452; NOTCH.
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 2.
 DR SMART: SM00004; NL; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50086; ANK_REPEAT; 3.
 DR PROSITE: PS00101; ASX_HYDROXYL; 3.
 DR PROSITE: PS00022; EGF_1; 12.
 DR PROSITE: PS01186; EGF_2; 11.
 DR PROSITE: PS50026; EGF_3; 13.
 DR PROSITE: PS01187; EGF_CA; 2.
 KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 1429 LIN-12 PROTEIN.
 FT DOMAIN 16 908 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 909 931 POTENTIAL.
 FT DOMAIN 932 1429 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 61 EGF-LIKE 1.
 FT DOMAIN 114 150 EGF-LIKE 2.
 FT DOMAIN 152 190 EGF-LIKE 3.
 FT DOMAIN 201 246 EGF-LIKE 4.
 FT DOMAIN 250 285 EGF-LIKE 5.
 FT DOMAIN 287 323 EGF-LIKE 6.
 FT DOMAIN 323 363 EGF-LIKE 7.

```

FT DOMAIN 365 402 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 404 441 EGF-LIKE 9.
FT DOMAIN 449 492 EGF-LIKE 10.
FT DOMAIN 503 541 EGF-LIKE 11.
FT DOMAIN 543 579 EGF-LIKE 12.
FT DOMAIN 582 619 EGF-LIKE 13.
FT REPEAT 635 669 EGF-LIKE 1.
FT REPEAT 670 710 LIN/NOTCH 2.
FT REPEAT 711 750 LIN/NOTCH 3.
FT REPEAT 1093 1122 ANK 1.
FT REPEAT 1126 1158 ANK 2.
FT REPEAT 1162 1194 ANK 3.
FT REPEAT 1206 1236 ANK 4.
FT REPEAT 1240 1269 ANK 5.
FT DISULFID 24 35 BY SIMILARITY.
FT DISULFID 29 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 118 129 BY SIMILARITY.
FT DISULFID 123 138 BY SIMILARITY.
FT DISULFID 140 149 BY SIMILARITY.
FT DISULFID 156 169 BY SIMILARITY.
FT DISULFID 163 178 BY SIMILARITY.
FT DISULFID 180 189 BY SIMILARITY.
FT DISULFID 205 237 BY SIMILARITY.
FT DISULFID 221 234 BY SIMILARITY.
FT DISULFID 236 245 BY SIMILARITY.
FT DISULFID 254 264 BY SIMILARITY.
FT DISULFID 259 273 BY SIMILARITY.
FT DISULFID 275 284 BY SIMILARITY.
FT DISULFID 291 302 BY SIMILARITY.
FT DISULFID 296 311 BY SIMILARITY.
FT DISULFID 313 322 BY SIMILARITY.
FT DISULFID 327 339 BY SIMILARITY.
FT DISULFID 334 351 BY SIMILARITY.
FT DISULFID 353 362 BY SIMILARITY.
FT DISULFID 369 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 429 BY SIMILARITY.
FT DISULFID 431 440 BY SIMILARITY.
FT DISULFID 440 458 BY SIMILARITY.
FT DISULFID 507 518 BY SIMILARITY.
FT DISULFID 512 529 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 547 558 BY SIMILARITY.
FT DISULFID 567 578 BY SIMILARITY.
FT DISULFID 586 597 BY SIMILARITY.
FT DISULFID 591 607 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 155 165 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 184 194 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 900 900 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 1429 AA; 157115 MW; 2558DD7A62C035DB CRC64;

```

Query Match 3.4%; Score 166.5; DB 1; Length 1429;
 Best Local Similarity 20.4%; Pred. No. 0.024;
 Matches 113; Conservative 43; Mismatches 172; Indels 225; Gaps 22;

```

QY 438 CAEGCGSGWIKGVCYCDKACNNACWMDGDDGSGNS---GSGRYTAGGGCTGSGVGHFWQ 494
DB 643 CSERA-----NDGNCDDACNCAACKFDGDCSGKREPKSKRY-----GNMCADF--- 687
QY 495 FGGGINSVSYCNQGCAN-----SWLADKFCDOACNVLSG 528
DB 688 FANGV-----CNQACNNEBCLYDGMDCLPAYVRCVPKIRHCASRFANGICDPECNNGC 742

```

```

QY 529 GPDAGCGQDPHFELHYKVLIPNOTHYIIPKGBCLPYFSAFAVAKRGVEGASDNPILRH 588
DB 743 GFDGGDC-----DNET----- 753
QY 589 ASIANKKKTIHLIMHSGMNAATTI-HPNLTFPQNTNDEFK-----NQITVEVD 634
DB 754 -----NATITINIRITVQ-MDPKEFCVTGGQSLMEISSALRVYRIQ 794
QY 635 -TRBPGKLNSTAQGVENLVSPITL-----LPEAFILPEDPKPKRPKFRHVNSTRRA 689
DB 795 RDEBGPV---FQWNGSEMDRVKMKERQITEGHVLISTISR-----KIKRSATNIGVVV 846
QY 690 QEEVKIPLVNISLPPDAQSLNTLDLOLEHGDTIKGYVLS-----KSALLRSFL- 740
DB 847 YLEFGNCDPDKCLYDQGVSDVSIARLAKKIDSFGLPISALVAERKSNNTGFLS 906
QY 741 -----MNSGHAKIKNOAITTDET 758
DB 907 WNAALLIGAGCLIVWVVLGALPGNRTKRRRMINASVMPMEWNEKRNKNQSISSQ 966
QY 759 NDSIVAPQERQV-----HKSILPNSLGVSF-----RLQRLTPPAVSQVNG 799
DB 967 HSLHLSYDGYIKQRNDELQHYSLYENPQSYGNGNDFLDPNHTNLIQITPEPESEPIKL 1026
QY 800 HDGQNPPLDLETTARFVE---THQKT---IGGNVTKEKPPSLIVPLESGMTREKKI 852
DB 1027 HTEAGSAYATEPITRESVNIIDPRHNRVYLHWIASNSAKKEDILV---HEACEIA 1082
QY 853 TGKEKENSMEEN 865
DB 1083 AGADVNMDCDEN 1095

```

```

RESULT 5
NTCL RAT STANDARD; PRT; 2511 AA.
AC Q07005;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
OS NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Schwann cell;
RX MEDLINE=9211383; PubMed=1764995;
RA Weinmaster G., Roberts V.J., Lemke G.;
RA "A homolog of Drosophila Notch expressed during mammalian
  development 113:199-205(1991).
  [2]
RP REVISIONS TO 1652-1653.
RA Weinmaster G.;
RA Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN FUNCTION.
RX MEDLINE=21094508; PubMed=1182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Honjo T.;
RA "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
  neural progenitor cells to an astroglial fate.";
  Neuron 29:45-55(2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=93202015; PubMed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RA "Notch2: a second mammalian Notch gene.";
  Development 116:931-941(1992).
RN [5]

```

RP TISSUE SPECIFICITY
 RX MEDLINE:21331789; PubMed:11438922;
 RA Irvn D.K., Zurcher S.D., Nguyen T., Weimaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 functional roles for the Notch-Delta signaling system during brain
 development.";
 RT J. Comp. Neurol. 436:167-181 (2001).
 RL
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 jagged2, jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released cell-intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
 CC Expressed in postnatal central nervous system (CNS) germinal zones
 CC and, in early postnatal life, within numerous cells throughout the
 CC CNS. Found in both subventricular and ventricular germinal zones.
 CC -1- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
 CC days 12 and 14 and decrease rapidly to much lower levels in the
 CC adult.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 36 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X57405; CAA40667.1; -
 DR HSSP: P00740; 1EDM.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl_S.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR006209; EGF-like.
 DR InterPro: IPR002949; Laminin_EGF.
 DR InterPro: IPR008287; Notch.
 DR InterPro: IPR000800; Notch_dom.
 DR Pfam: PF00023; ank; 6.
 DR Pfam: PF00066; notch; 3.
 DR Pfam: PF000279; Notch; 1.
 DR PRINTS: PR00010; EGFLOOD.
 DR PRINTS: PR00011; EGFAMININ.
 DR PRINTS: PR01452; NOTCH.

DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39 99
 DR DOMAIN 102 139
 DR DOMAIN 140 176
 DR DOMAIN 178 216
 DR DOMAIN 218 255
 DR DOMAIN 257 293
 DR DOMAIN 295 333
 DR DOMAIN 335 371
 DR DOMAIN 372 410
 DR DOMAIN 412 450
 DR DOMAIN 452 488
 DR DOMAIN 490 526
 DR DOMAIN 528 564
 DR DOMAIN 566 601
 DR DOMAIN 603 639
 DR DOMAIN 641 676
 DR DOMAIN 678 714
 DR DOMAIN 716 751
 DR DOMAIN 753 789
 DR DOMAIN 791 827
 DR DOMAIN 829 867
 DR DOMAIN 869 905
 DR DOMAIN 907 943
 DR DOMAIN 945 981
 DR DOMAIN 1021 1057
 DR DOMAIN 1059 1095
 DR DOMAIN 1097 1143
 DR DOMAIN 1145 1181
 DR DOMAIN 1183 1219
 DR DOMAIN 1221 1265
 DR DOMAIN 1267 1305
 DR DOMAIN 1307 1346
 DR DOMAIN 1348 1384
 DR DOMAIN 1387 1426
 DR DOMAIN 1445 1480
 DR DOMAIN 1481 1522
 DR DOMAIN 1523 1562
 DR DOMAIN 1917 1946
 DR DOMAIN 1950 1980
 DR DOMAIN 1984 2013
 DR DOMAIN 2017 2046
 DR DOMAIN 2050 2079
 DR DOMAIN 1730 1733
 DR DOMAIN 1891 1894
 DR DOMAIN 2258 2261
 DR DOMAIN 2497 2500
 DR SITE 1654 1655
 DR DISULFID 24 37
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00179; EGF_CA; 25.
 DR SMART: SM00004; NU; 2.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 35.
 DR PROSITE: PS0186; EGF_2; 26.
 DR PROSITE: PS50026; EGF_3; 36.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR PROSITE: PS0187; EGF_Ca; 21.
 DR Receptor: Transcription regulation; Activator; Differentiation;
 DR Receptor: Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 DR SIGNAL 1 18
 DR CHAIN 19 2531
 DR CHAIN 1711 2531
 DR CHAIN 1744 2531
 DR CHAIN 19 1723
 DR TRANSMEM 1724 1746
 DR DOMAIN 1747 2531
 DR DOMAIN 20 58
 DR DOMAIN 39

FT	DISULFID	31	46		BY SIMILARITY.
FT	DISULFID	48	57		BY SIMILARITY.
FT	DISULFID	63	74		BY SIMILARITY.
FT	DISULFID	68	87		BY SIMILARITY.
FT	DISULFID	89	98		BY SIMILARITY.
FT	DISULFID	106	117		BY SIMILARITY.
FT	DISULFID	111	127		BY SIMILARITY.
FT	DISULFID	129	138		BY SIMILARITY.
FT	DISULFID	144	155		BY SIMILARITY.
FT	DISULFID	149	164		BY SIMILARITY.
FT	DISULFID	166	175		BY SIMILARITY.
FT	DISULFID	182	195		BY SIMILARITY.
FT	DISULFID	189	204		BY SIMILARITY.
FT	DISULFID	206	215		BY SIMILARITY.
FT	DISULFID	222	233		BY SIMILARITY.
FT	DISULFID	227	243		BY SIMILARITY.
Query Match 3.4%; Score 165; DB 1; Length 2531;					
Best Local Similarity 20.6%; Pred. No. 0.072;					
Matches 111; Conservative 63; Mismatches 179; Indels 186; Gaps 27;					
OY	412	GKDWPPDPFSHSGQKVYLTPVPCNACBGPSSWIMDYGCDAACNNASCDMDGPGCSGN	471		
Dh	1437	GRDLP-----PQIEFACEIPCEQEDCAKNK-----VCNIQCNNHAGCWDGGSCSIN	1482		
OY	472	SGGS-----RYIAGG-----GATGSIGVHPMFGGINSVSYSN-----QGCA	510		
Dh	1483	FNDWKNTQSLOQCMKFSDGHCHDSQCNSAGCLFDGFDCQLTBSG-----QCNPLOYCYCK	1537		
OY	511	NSWLADFCPOACNVLSGCFDAGCGODHFHE-----LYKVLLP-----NOTHYITP	558		
Dh	1538	DHF-SDGHCQGCNSABCEWDGLDCA-EHVPERLAAGTLVLVLLPPDQLRNNSFFHLDD	1595		
OY	559	KGECL-----PYFSFAEVAKRVEGAYSDNPILIRHASTANKMTIHDI	601		
Dh	1596	VSHVLHNIVVPFKRAOQGMIFPV-----GREELRKHIHKISAV---CWATTSL	1644		
OY	602	-----MHSGMAHTTHFNLTFOINTDEEPKQIIVEVDTRZ-----GP	639		
Dh	1645	PGTNGRHORELDTMDHGHSIVY-----LEIDNRCCVSSSQCFOSATDVAA	1691		
OY	640	KLNSTAQGYENLVSPITLLPEAEILFPEDIP-----KE	672		
Dh	1692	PLGLALSLGSINITYKLEAV-KSETVPRPRLPSQLHMYAAAAPULLFTVCGVLLSRKR	1750		
OY	673	KR-----PK-FKRDVNSTRAQEVEVKIPL-VNISLLP-KDAQLSINTLDJLEHG-	721		
Dh	1751	RROHQLMFPEGFKVSBASKKRR-----FLGEDSVGLKPLKNASGALMDNQEMWD	1805		
OY	722	-DILTKVYNLSKSLIASFLMNSOHAKIKNOAITTDETDSLVAPOEKQVHKSILNSLG	780		
Dh	1806	EDLTETKFRFEPEFVLDDLDQDTHRWTOQHLDADLRVSAAPPTPQQ-----	1854		
OY	781	VSERLQRLTEPAVSVKNGHDQGNPRLDETARFEVETHQTIGCVNKYEKPSSLI	839		
Dh	1855	-----GEVVDADCMMVNVRGPD--GFTPLMTASCSCGGLER-----GNSEBEDARA VI	1900		
RESULT 6					
NCBI_MOUSE STANDARD; PRT; 2531 AA.					
ID	NCBI_MOUSE	001705; Q06007; Q61905; Q99JC3; Q9QW58; Q9R0X7;			
AC	001705; Q06007; Q61905; Q99JC3; Q9QW58; Q9R0X7;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Moch A)				
DE	(mrl4) (p300).				
GN	NOTCH1 OR NOTCH.				
OC	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Mol.				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
NX	NCBI_TaxId=10090;				
NX	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RC	TISSUE=Embryo;				
RX	MEDLINE=93194170; PubMed=8449489;				
RX	Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,				
RA	Copeland N.G., Gridley T.;				
RT	"Cloning, analysis, and chromosomal localization of Notch-1, a mouse				
RT	homolog of Drosophila Notch."				
RL	Genomics 15:259-264(1993).				
RP	SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.				
RC	STRAIN=CD-1; TISSUE=Embryo;				
RX	MEDLINE=93050801; PubMed=1426644;				
RA	Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.;				
RT	"Expression analysis of a Notch homologue in the mouse embryo."				
RL	Dev. Biol. 154:377-387(1992).				
RP	SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.				
RC	TISSUE=Embryo;				
RX	MEDLINE=93048935; PubMed=1425352;				
RA	Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,				
RA	Greenspan R.J., McMahon A.P., Gridley T.;				
RT	"Expression pattern of Notch, a mouse homolog of Drosophila Notch,				
RT	suggests an important role in early postimplantation mouse				
RL	development 115:737-744(1992).				
RP	SEQUENCE OF 1161-1547 FROM N.A.				
RC	STRAIN=C57BL/6 X CBA; TISSUE=embryo;				
RX	MEDLINE=93170563; PubMed=8440332;				
RA	Lardelli M., Lendahl U.;				
RT	"Moch A and Notch B two mouse Notch homologues coexpressed in a				
RT	wide variety of tissues."				
RL	Exp. Cell Res. 204:364-372(1993).				
RP	SEQUENCE OF 1659-1673 FROM N.A.				
RC	MEDLINE=99364499; PubMed=10437789;				
RX	Lee U.S., Ishimoto A., Yanagawa S.T.;				
RT	"Mune leukemia provirus-mediated activation of the Notch1 gene leads				
RT	to induction of HES-1 in a mouse T lymphoma cell line, DL-3."				
RL	FEBS Lett. 455:276-280(1999).				
RP	SEQUENCE OF 1950-2201 FROM N.A.				
RC	MEDLINE=98029496; PubMed=9384671;				
RX	Messierie M., Folio M., Neltis M., Eggert H., Boehm T.;				
RT	"Dynamic changes in gene expression during in vitro differentiation of				
RT	mouse embryonic stem cells."				
RL	Cytokines Cell. Mol. Ther. 1:139-143(1995).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F., Beesla C., Brou C., LeBail O., Jarrault S., Seidah N.G.,				
RA	Israel A.;				
RT	"The Notch1 receptor is cleaved constitutively by a furin-like				
RT	convertase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).				
RP	SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND				
RP	MUTAGENESIS OF 1651-ARG-ARG-1654.				
RX	MEDLINE=98318619; PubMed=953148;				
RA	Logeat F				

RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
RA Okano H., Matsuno K.,
RT "Murine homologs of *deltex* define a novel gene family involved in
RT vertebrate Notch signaling and neurogenesis." ;
RT Int. J. Dev. Neurosci. 19:21-35 (2001)
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Vagued, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation. May be involved in mesoderm
CC development, somite formation and neurogenesis. Involved in the
CC maturation of both CD4+ and CD8+ cells in the thymus.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(IEC) which are probably linked by disulfide
CC bonds. Interacts with DTX1 and DTX2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC Name=2;
CC IsoId=Q01705-1; Sequence=Displayed;
CC IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
CC thymus. Expressed at lower levels in the spleen, bone-marrow,
CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
CC muscle, kidney and heart.
CC -1- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
CC endothelial cells, while much lower levels are seen in the
CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
CC the neuroepithelium. At 13.5 dpc expressed in the surface
CC ectoderm, eye and developing whisker follicles.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(IEC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXTr). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z11886; CAA7941.1; -
CC EMBL; L02613; AAK14898.1; -
CC EMBL; X68278; CAA48339.1; -
CC EMBL; AJ238029; CAB40733.1; -
CC EMBL; X82562; CAA57909.1; -
CC F01; A46019; A46019.
CC F01; B49175; B49175.
CC HSSP; P00740; IEDM.
CC MGI; MGI:97363; Notch1.

DR GO; GO:0005887; C: integral to plasma membrane; IC.
DR GO; GO:0005515; F: protein binding; IP.
DR GO; GO:0030154; P: cell differentiation; IMP.
DR GO; GO:0007386; P: compartment specification; IMP.
DR GO; GO:0007219; P: N signaling pathway; IC.
DR GO; GO:0045944; P: positive regulation of transcription from P...; IDA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 7.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF02279; Notch; 1.
DR PRINTS; PRO0010; EGFLEOOD.
DR PRINTS; PRO01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS50026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KM Receptor; Transcription regulation; Activator; Differentiation;
KM Developmental protein; Repeat; ANK repeat; EGF-like domain;
KM Transmembrane; Glycoprotein; Signal; Phosphorylation;
KM Alternative splicing.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT CHAIN 1711 2531 NOTCH EXTRACELLULAR TRUNCATION.
FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN.
FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).
Query Match 3.3%; Score 163; DB 1; Length 2531;
Best Local Similarity 20.4%; Pred. No. 0.096;
Matches 119; Conservative 68; Mismatches 204; Indels 192; Gaps 29;
QY 434 PVENCAGC--PSGWIKDG--YCDKACNNACDMDGDCSGSGS-----RYI 478
DB 1441 PPGQIEACELPEGCYDAENKVCNLCNNHACGMGDCGDSLFNDPWKXCTGSLCKYKF 1500
QY 479 AGG-----GGTSGIVGHPWQFGGINSVSYCN-----QGANSLADKFCDOACNVLSG 528
DB 1501 SDHCSDQCNAGCLFDGFCDCLTED-----QCNPLYDQYCHDF-SDHCDCGQNSABC 1554
QY 529 GPDAGPCGDHFE-----LYKVILP-----NOTHYIIRKGLCPYFSPAYAKKGV 576
DB 1555 EWDGLDCA-EHPPERLAAGTLVAVLPDQIRNNSFHL--RELSHVLTNNVFKKDA 1610
QY 577 EGA5DNPIIRHNSAMK-----WKTILH-----MHGMMATTHFNLTFOFT 620
DB 1611 OGQGMIFPYGHEERKRPKIRKSTVGWATSSLLTSGSGRQRRELDPMIDIGSTIVY--- 1667
QY 621 NDEEFMQLTVETDTE-----GPKLNSAQGYNNVSPILLBEATL 665
DB 1668 -----LETDNQCVQSSSQCFQSATDVAAFGALASLNIPIKIAV--KSEPV 1716
QY 666 FEDIP-----KEK-----FPK-FKHVDNSTRRAOE 691
DB 1717 EPPLEPQLHMYAAAFVLPFVGGVLLSKRRRQGLWFPPEGFVSEASKKKRR- 1775
QY 692 EVKIPV--VNISLP-KDAQLSLNTDLENG--DITIKGNLSKSLRLBFLNNSQA 746
DB 1776 ----PLGSDSVGLKPLKNSDGLMDNDQNEWGEDLETKKRFEEBPVLPDLSQDTDR 1831

```

QY 747 KIKNOAIIITDETNDLSVAPQEQVHNSILNSLGVSERLQRLTPFAVSQVKNHGQGNP 806
DB 1832 QWTOQLDADADJMSMAFLTPQ-----GEVDADQMDVNVAGPD--GFT 1873
QY 807 PDLLETTARFVETHQKIGNVTKERKPSLI-----VLEQOMTKKKIT----- 853
DB 1874 PLMTASCSGGGLT-----GNSEEDAPAVISDFITYOGASLHQTRTGTLALHAR 1927
QY 854 -GKEKENSMEENAEENH-----GTEVLL 877
DB 1928 YSRSDRKRRLBASADANIQDNGRTPLHAASADAGVQILL 1970

RESULT 7
NTCL BRARE STANDARD; PRT; 2437 AA.
AC P46530;
DT 01-NOV-1995 (Rel. 32, Last Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor.
GN NOTCH1 OR NOTCH.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxId=7955;
RN RP
RP [1]
RC TISSUE=Embryo;
RC MEDLINE=94128602; PubMed=8297791;
RA Bierkamp C., Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RL its pattern of transcription during early embryogenesis.";
RL Mech. Dev. 43:87-100(1993).
CC - FUNCTION: Implicated in cell fate specifications during
CC embryonic development. May be involved in the formation of the
CC neural plate, notochord and brain vesicles.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
CC stages. During gastrulation is differentially expressed,
CC accumulating predominantly in the prechordal mesoderm and
CC notochord. At the end of gastrulation, expressed along the
CC anterior-posterior axis including the developing neural plate
CC and differentiating mesoderm. Also present in the developing
CC brain and head regions.
CC - SIMILARITY: Belongs to the NOTCH family.
CC - SIMILARITY: Contains 36 EGF-like domains.
CC - SIMILARITY: Contains 3 Lin/Notch repeats.
CC - SIMILARITY: Contains 6 ANK repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, X69088; CAA48631.1; -
DR PIR, S42612; S42612.
DR HSSP, P00740; 1EDM.
DR ZFIN, ZDB-GENE-990415-173; notch1a.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001981; EGF_Ca.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00065; notch; 3.
DR PIRSF; PIRSF002379; Notch; 1.
DR PRINTS; PR00010; EGFBLD.
DR PRINTS; PR00011; EGF-LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 22.
DR SMART; SM00004; NU; 3.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 28.
DR PROSITE; PS50026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 22.
DR Receptor; Transcription regulation; Activator; Differentiation;
KM Developmental protein; Neurogenesis; Repeat; ANK repeat;
KT EGF-like domain; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 787
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT REPEAT 1446 1486
FT REPEAT 1487 1520
FT REPEAT 1521 1561
FT REPEAT 1867 1910
FT REPEAT 1915 1944
FT REPEAT 1948 1978
FT REPEAT 1982 2011
FT REPEAT 2015 2044
FT REPEAT 2048 2077
FT DOMAIN 2265 2276
FT SITE 1656 1657

```

FT	DISULFID	25	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	73	BY SIMILARITY.
FT	DISULFID	67	86	BY SIMILARITY.
FT	DISULFID	88	97	BY SIMILARITY.
FT	DISULFID	105	116	BY SIMILARITY.
FT	DISULFID	110	126	BY SIMILARITY.
FT	DISULFID	128	137	BY SIMILARITY.
FT	DISULFID	143	154	BY SIMILARITY.
FT	DISULFID	148	163	BY SIMILARITY.
FT	DISULFID	165	174	BY SIMILARITY.
FT	DISULFID	181	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	465	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	503	BY SIMILARITY.
FT	DISULFID	498	512	BY SIMILARITY.
FT	DISULFID	514	523	BY SIMILARITY.
FT	DISULFID	530	541	BY SIMILARITY.
FT	DISULFID	535	550	BY SIMILARITY.
FT	DISULFID	552	561	BY SIMILARITY.
FT	DISULFID	568	578	BY SIMILARITY.
FT	DISULFID	573	587	BY SIMILARITY.
FT	DISULFID	589	598	BY SIMILARITY.
FT	DISULFID	605	616	BY SIMILARITY.
FT	DISULFID	610	625	BY SIMILARITY.
FT	DISULFID	627	636	BY SIMILARITY.
FT	DISULFID	643	653	BY SIMILARITY.
FT	DISULFID	648	662	BY SIMILARITY.
FT	DISULFID	664	673	BY SIMILARITY.
FT	DISULFID	680	691	BY SIMILARITY.
FT	DISULFID	685	700	BY SIMILARITY.
FT	DISULFID	702	711	BY SIMILARITY.
FT	DISULFID	718	728	BY SIMILARITY.
FT	DISULFID	723	737	BY SIMILARITY.
FT	DISULFID	739	748	BY SIMILARITY.
FT	DISULFID	755	766	BY SIMILARITY.
FT	DISULFID	760	775	BY SIMILARITY.
FT	DISULFID	777	786	BY SIMILARITY.
FT	DISULFID	793	804	BY SIMILARITY.
FT	DISULFID	798	813	BY SIMILARITY.
FT	DISULFID	815	824	BY SIMILARITY.
FT	DISULFID	831	842	BY SIMILARITY.
FT	DISULFID	836	853	BY SIMILARITY.
FT	DISULFID	855	864	BY SIMILARITY.
FT	DISULFID	871	882	BY SIMILARITY.
FT	DISULFID	876	891	BY SIMILARITY.
FT	DISULFID	893	902	BY SIMILARITY.
FT	DISULFID	909	920	BY SIMILARITY.

Query Match 3.3%; Score 162.5; DB 1; Length 2437;
Best Local Similarity 20.7%; Pred. No. 0.097;

Matches 119; Conservative 58; Mismatches 201; Indels 197; Gaps 29;	
QY 421 YSHSKQ-----KVYLTPVPNCAGCPGSMWKDGYCDACNNSACMDGDCSGNS	472
Db 1427 YSPSGGQGDIAPEVVEIRCEIAQC-EGRGNAI-----CDTCNNHAGCGDGC----	1477
QY 473 GGSRYIAGGGGGSIVGHPMQ-----FGGINSVSYCNOGANS-----	512
Db 1478 -----SINFPDPMQCSAALOCMRYFNDG-----KCDQCCTACCLYDGFDC	1519
QY 513 -----WLDKFCDCQACNVLSGCFDAGCGOD-----HFHLYVYILP	550
Db 1520 QRLBGCQNLVYDQYCDHVDADHCCDGCNNACEMDGLDADVDQKLAVSLVYVHP	1579
QY 551 -----NQTNYII-----PKGECI-PIYSPFAVAKGVGAYSDNPITR	587
Db 1580 PDELRRSSFPRESLSLHTNVFRRDANGALLFPYGG-----SSEHLSK	1626
QY 588 HASIAKMTIHLIMSGNATTIHPN-LTFONTNDEEPMQIT-VEVDTR-----	637
Db 1627 HK-REDWIDPQGLMORARSLTSPFKPTRELDHMEVKSIVYLEIDNRCCFOQSDC	1684
QY 638 -----GPKINSTAGXYENLVSPITLLPEABII-FEDIPKRPKPKFRHDVNSTRAQ	690
Db 1685 FQSATDVAAPLGLASSG--NINVP-----YIIEAVTSRSGPSPK-----TG	1723
QY 691 EEVKIPLVNIILPKDAQLSMTLTL-----OLEHGDITLKGYNLSKSLRFLMNSQ	744
Db 1724 EMTFPMFLVLLAL-----AVLALAAGVYVSRKKRHHGQIMFP-----EGKRVNEP	1769
QY 745 HAKIKNOAIIITETDNDLVAPOEKOVHKSILPNSI---GVSERLQRLFPNAVSV-KVNG	799
Db 1770 KKKRREPV-----GEDSVGLKPLKNSDSLMDQSEMAEDTNRKFRFGSILMSGQ	1824
QY 800 --HDQGNPPLDLETTAFRFRVETHQKTIGAVTKE-----KPSLIYPLSOMTKEXK	851
Db 1825 LDRHQWTOQHLD---AADRLNSMAFPPQGGIENDCMQDYNVYRGDFGFTPL---MIA	1878
QY 852 ITSEKENSREMEENHIGTEVLLGRKLCQHYTD	886
Db 1879 GGGLNENGBABEDPSADYITDPIYHGANLHNQTD	1913
RESULT 8	
ID NTG3 HUMAN STANDARD; PRT; 2321 AA.	
AC Q9UM47: Q9UEB3: Q9UPL3: Q9Y6L8;	
DT 28-FEB-2003 (Rel. 41, Created)	
DT 28-FEB-2003 (Rel. 41, Last sequence update)	
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).	
GN NOTCH3.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX NCBI_TaxID=9606;	
RN (1)	
RP SEQUENCE FROM N.A.	
RX MEDLINE=97032728; PubMed=8878478;	
RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,	
RA Alamowitch S., Domenga V., Cecillon M., Mareschal E., Maciazek J.,	
RA Vayssières C., Craud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,	
RA Bach J.-F., Bousset M.-G., Tournier-Lasserre E.,	
RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition	
RT causing stroke and dementia.";	
RL Nature 383:707-710(1996).	
RN (2)	
RP SEQUENCE FROM N.A.	
RA Guinel M., Artavanis-Tsakonas S.;	
RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.	
RN (3)	
RP SEQUENCE FROM N.A.	
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.;	

RA Burkhardt-Schulz K., Gordon L., Kyle A., Ramirez M., Stiliwagen S.,
 RA Phan H., Velasco N., Gannes D., Dangman L., Poundstone P.,
 RA Christensen M., Georgescu A., Ayila U., Liu S., Altix C., Andreise T.,
 RA Trankelm M., Ambro-Keller G., Coetfield J., Duarte S., Lucas S.,
 RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
 RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
 RA Carraro A.V.,
 RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFRL) cluster in
 RT 19p13.1.",
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [4].
 RP VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
 RP ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
 RP GLY-222; TYR-242; CYS-258; TYR-552; CYS-558; CYS-578; CYS-728;
 RP CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
 RP ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
 RX MEDLINE=98049753; PubMed=9388399;
 RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
 RA Vayssiere C., Cruaud C., Macchatek J., Weissensbach J., Bousser M.-G.,
 RA Bach J.-F., Tournier-Lasserre E.,
 RT "Strong clustering and stereotyped nature of Notch3 mutations in
 RT CADASIL patients.",
 RL Lancet 350:1511-1515 (1997).
 RN [5].
 RP VARIANT CADASIL 114-GLY -PRO-120 DEL.
 RX MEDLINE=20264473; PubMed=10802807;
 RA Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,
 RA Rouchoux M.M., Lucas C., Lays D., Bousser M.-G., Tournier-Lasserre E.,
 RT "Splice site mutation causing a seven amino acid Notch3 in-frame
 RT deletion in CADASIL.",
 RL Neurology 54:1874-1875 (2000).
 RN [6].
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Garcanguil M.-L.,
 RA Banks A., Leisman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.,
 RL Am. J. Pathol. 154:785-794 (1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-1/ kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (by similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (by similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
 CC tissues.
 CC -1- PM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (by similarity).
 CC -1- PM: Phosphorylated (by similarity).
 CC -1- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal
 CC dominant arteriopathy with subcortical infarcts and
 CC leukoencephalopathy (CADASIL) [MIM:15310]. CADASIL causes a type
 CC of stroke and dementia of which key features include recurrent
 CC subcortical ischemic events and vascular dementia.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 34 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.

	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed.
	CC	User fee by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/termsandconditions or send an email to license@ebi.ac.uk).
	CC	
	DR	EMBL; U97669; AAB91377.1; --
	DR	EMBL; AF058890; AAC14346.1; JOINED.
	DR	EMBL; AF058881; AAC14346.1; JOINED.
	DR	EMBL; AF058882; AAC14346.1; JOINED.
	DR	EMBL; AF058883; AAC14346.1; JOINED.
	DR	EMBL; AF058884; AAC14346.1; JOINED.
	DR	EMBL; AF058885; AAC14346.1; JOINED.
	DR	EMBL; AF058886; AAC14346.1; JOINED.
	DR	EMBL; AF058887; AAC14346.1; JOINED.
	DR	EMBL; AF058888; AAC14346.1; JOINED.
	DR	EMBL; AF058889; AAC14346.1; JOINED.
	DR	EMBL; AF058890; AAC14346.1; JOINED.
	DR	EMBL; AF058891; AAC14346.1; JOINED.
	DR	EMBL; AF058892; AAC14346.1; JOINED.
	DR	EMBL; AF058893; AAC14346.1; JOINED.
	DR	EMBL; AF058894; AAC14346.1; JOINED.
	DR	EMBL; AF058895; AAC14346.1; JOINED.
	DR	EMBL; AF058896; AAC14346.1; JOINED.
	DR	EMBL; AF058897; AAC14346.1; JOINED.
	DR	EMBL; AF058898; AAC14346.1; JOINED.
	DR	EMBL; AF058899; AAC14346.1; JOINED.
	DR	EMBL; AC004257; AAC04897.1; --
	DR	EMBL; AC004653; AAC15789.1; ALT_INIT.
	DR	PIR; S78549; S78549.
	DR	HSSP; P00740; IEDM.
	DR	Genev; HGNC:7883; NOTCH3.
	DR	MIM; 600276; --
	DR	MIM; 125310; --
	DR	InterPro; IPRO02110; ANK.
	DR	InterPro; IPRO00152; Asx_hydroxy_1_S.
	DR	InterPro; IPRO00742; EGF_2.
	DR	InterPro; IPRO01881; EGF_CA.
	DR	InterPro; IPRO01438; EGF_T1.
	DR	InterPro; IPRO06209; EGF_Like.
	DR	InterPro; IPRO02049; Laminin_EGF.
	DR	InterPro; IPRO08297; Notch.
	DR	InterPro; IPRO00800; Notch_dom.
	DR	pfam; PF00023; ank; 5.
	DR	pfam; PR00008; EGF; 34.
	DR	pfam; PR00066; notch; 3.
	DR	PIRSF; PIRSFO02279; Notch; 1.
	DR	PRINTS; PRO0010; EGFLWOOD.
	DR	PRINTS; PRO0011; EGFLAMININ.
	DR	PRINTS; PRO1452; NOTCH.
	DR	SMART; SMART00248; ANK; 6.
	DR	SMART; SMART00179; EGF_CA; 19.
	DR	SMART; SMART00064; NL; 3.
	DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.
	DR	PROSITE; PS50088; ANK_REPEAT; 4.
	DR	PROSITE; PS00010; ASX_HYDROXYL; 18.
	DR	PROSITE; PS00022; EGF_1; 33.
	DR	PROSITE; PS00186; EGF_2; 25.
	DR	PROSITE; PS50026; EGF_3; 34.
	DR	PROSITE; PS01187; EGF_CA; 16.
	KW	Receptor; Transcription regulation; Activator; Differentiation;
	KW	Developmental protein; Repeat; ANK repeat; EGF-like domain;
	KW	Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
	KW	Disease mutation.
	FT	SIGNAL 1..39 POTENTIAL.
	FT	CHAIN 40..2321 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
	FT	CHAIN 1629..2321 NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
	FT	CHAIN 1662..2321 NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY).
	TT	


```

CC      from the membrane.
CC      -!- PTM: Phosphorylated.
CC      -!- DISEASE: Loss of the extracellular domain causes constitutive
CC      activation of the Notch protein, which leads to hyperproliferation
CC      of gladiular epithelial tissues and development of mammary
CC      carcinomas.
CC      -!- SIMILARITY: Belongs to the NOTCH family.
CC      -!- SIMILARITY: Contains 29 EGF-like domains.
CC      -!- SIMILARITY: Contains 3 Lrr/Notch repeats.
CC      -!- SIMILARITY: Contains 5 ANK repeats.
CC      -----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see license@isb-sib.ch).
CC      -----
CC      EMBL, M80456; AAB38377.1; -.
CC      EMBL, U43691; AAC52630.1; -.
CC      EMBL, U43691; AAC52631.1; -.
CC      EMBL, AF030001; AAB82004.1; -.
CC      EMBL, AB016771; BAA32281.1; ALT_SEQ.
CC      EMBL, AB016772; BAA32283.1; ALT_INIT.
CC      EMBL, AB016773; BAA32284.1; ALT_INIT.
CC      EMBL, AB016774; BAA32285.1; -.
CC      PIR, A38072; TWMVT3.
CC      PIR, T09059; T09059.
CC      HSSP, P08709; LBR9.
CC      MGD, MGI:107471; Notch4.
CC      InterPro, IPR002110; ANK.
CC      InterPro, IPR000152; Asx_hydroxyl_5.
CC      InterPro, IPR000742; EGF_2.
CC      InterPro, IPR001881; EGF_CA.
CC      InterPro, IPR001438; EGF_II.
CC      InterPro, IPR006209; EGF_III.
CC      InterPro, IPR002049; Laminin_EGF.
CC      InterPro, IPR008297; Notch.
CC      InterPro, IPR008000; Notch_dom.
CC      Pfam, PF00023; ank; 6.
CC      Pfam, PF00008; EGF; 27.
CC      Pfam, PF00065; notch; 2.
CC      Pfam, PF0002279; Notch; 1.
CC      PRINTS, PIRSF002279; Notch; 1.
CC      PRINTS, PRK0010; EGFBLOOD.
CC      PRINTS, PRK0011; EGFLAMININ.
CC      PRINTS, PRO1452; NOTCH.
CC      SMART, SMO0248; ANK; 6.
CC      SMART, SMO0179; EGF_CA; 11.
CC      SMART, SMO0004; NL; 2.
CC      PROSITE, PS50297; ANK_REPEAT_REGION; 1.
CC      PROSITE, PS50088; ANK_REPEAT; 5.
CC      PROSITE, PS00010; ASX_HYDROXYL; 11.
CC      PROSITE, PS00022; EGF_1; 28.
CC      PROSITE, PS01186; EGF_2; 21.
CC      PROSITE, PS50026; EGF_3; 27.
CC      PROSITE, PS01187; EGF_CA; 9.
CC      Receptor, Transcription regulation; Activator; Differentiation;
CC      Developmental protein; Repeat; ANK repeat; EGF-like domain;
CC      Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
CC      SIGNAL 1 20
CC      CHAIN 1 11 1964
CC      CHAIN 1411 1964
CC      CHAIN 1428 1964
CC      CHAIN 1463 1964
CC      CHAIN 21 1443
CC      TRANSMEM 1444 1464
CC      TRANSMEM 1465 1964
CC      DOMAIN 21 60
CC      DOMAIN 61 112
CC      DOMAIN 115 152
CC      DOMAIN 153 189
CC      DOMAIN 191 229
CC      EGF-LIKE 1.
CC      EGF-LIKE 2.
CC      EGF-LIKE 3.
CC      EGF-LIKE 4.
CC      EGF-LIKE 5.
CC      CALCIUM-BINDING (POTENTIAL).

```

Query Match	Best Local Match	Similarity	Score	DB 1	Length
Matches	41	Conservative 11; Mismatches 32; Indels 62; Gaps 8;	28.1%; Pred. No. 0.32;		
FT	DOMAIN	273	271	EGF-LIKE 6.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	273	309	EGF-LIKE 7.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	311	350	EGF-LIKE 8.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	352	388	EGF-LIKE 9.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	389	427	EGF-LIKE 10.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	429	508	EGF-LIKE 11.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	472	546	EGF-LIKE 12.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	510	584	EGF-LIKE 13.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	548	584	EGF-LIKE 14.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	586	622	EGF-LIKE 15.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	623	656	EGF-LIKE 16.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	658	686	EGF-LIKE 17.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	688	724	EGF-LIKE 18.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	726	762	EGF-LIKE 19.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	764	800	EGF-LIKE 20.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	803	839	EGF-LIKE 21.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	841	877	EGF-LIKE 22.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	878	924	EGF-LIKE 23.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	926	962	EGF-LIKE 24.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	964	1000	EGF-LIKE 25.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1002	1040	EGF-LIKE 26.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1042	1081	EGF-LIKE 27.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1083	1122	EGF-LIKE 28.	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1126	1167	EGF-LIKE 29.	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1168	1208	EGF-LIKE 30.	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1209	1242	LIN/NOTCH 1.	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1243	1282	LIN/NOTCH 2.	CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1243	1282	LIN/NOTCH 3.	CALCIUM-BINDING (POTENTIAL)
Query Match	Best Local Match	Similarity	Score	DB 1	Length
Matches	41	Conservative 11; Mismatches 32; Indels 62; Gaps 8;	28.1%; Pred. No. 0.32;		
QY	434	PVPNC-----AEGCGSMIKQYCYCKACNNASCDMDGDCGSGSGRYIAGGCTGSGV	489		
DB	1162	PGPRCQGRPGASGGCGRG-RGDTCDAGSGSGPGMDGDC-----SIGV	1203		
QY	490	GHPWQFGGGINSVYC-----NQGCA-----NSWLAKF	518		
DB	1204	PDPWK-----GGPRHQCWILFRDGRCHPQCSEELFDYGCERPLTCTPAYDYCRDHFH	1260		
QY	519	---CDQACNVLSGCFDAGDC---GOD	538		
DB	1261	NGHCEKGCNNACGMDGDCRPSSED	1286		
RESULT 10					
NTC4_HUMAN					
ID	NTC4_HUMAN	STANDARD;	PRT: 2003 AA.		
AC	O99466; O000306; Q99458; Q99940; Q9H588; Q9U119; Q9U10;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Neurogenic locus notch homolog protein 4 precursor (Notch 4)				
DE	(Notch4).				
OS	NOTCH4.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_Taxid:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.				
RC	TISSUE=Placenta;				
RC	MEDLINE=97311416; PubMed=9168133;				
RA	Sugaya K., Sasanuma S.-I., Nohata J., Kimura T., Fukagawa T.,				
RA	Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.,				
RT	"Gene organization of human NOTCH4 and (CTG)n polymorphism in this				
RL	gene. 189:1235-244(1997)."				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3).				
RC	TISSUE=Bone marrow, and Heart;				
RC	MEDLINE=98360091; PubMed=9693032;				
RA	Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,				

RA Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA
 sequence of the human NOTCH4 gene.";
 RL Genomics 51:45-58(1998).
 RN [3]
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
 RA Miyagawa T., Tokunaga K., Hojho H.;
 RT "Human notch4 gene variant.";
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis R., Carcangu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Aravanis-Taskonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(IEC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=Q99466-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q99466-2; Sequence=VSP_001406;
 CC Name=3;
 CC IsoId=Q99466-3; Sequence=VSP_001407;
 CC -1- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(IEC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- POLYMORPHISM: The poly-leu region of NOTCH4 (in the signal
 CC peptide) is polymorphic and the number of leu varies in the
 CC population (from 6 to 12).
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 28 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in position 1438 to 1463.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; D63395; BAA09708.1; ALT_FRAME.
 DR EMBL; D6566; BAA13116.1; -
 DR EMBL; U95299; AAC32288.1; -
 DR EMBL; U98335; AAC63097.1; -
 DR EMBL; AB023961; BAB20317.1; -
 DR EMBL; AB024520; BAA88951.1; -
 DR EMBL; AB024578; BAA88952.1; -
 DR HSP; P08709; IEP9.
 DR Genew; HGNC:7884; NOTCH4.
 DR MIM; 164951; -
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_5.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; LamAin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR008000; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF_26.
 DR Pfam; PF00066; notch_2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLDOD.
 DR PRINTS; PR00011; EGFFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00004; NU_3.
 DR PROSITE; PS50257; ANK_REPEAT; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS00010; ASX_HYDROXYL; 11.
 DR PROSITE; PS00022; EGF_1; 28.
 DR PROSITE; PS01186; EGF_2; 21.
 DR PROSITE; PS50026; EGF_3; 28.
 DR PROSITE; PS01187; EGF_CA; 9.
 KW Receptor; transcription regulation; Activator; Differentiation;
 KW Developmental protein; Protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Triplet repeat expansion; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 804 842
 FT DOMAIN 846 880
 FT DOMAIN 882 928
 FT CHAIN 1467 2003
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 116 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511

```
FT DOMAIN 930 966 EGF-LIKE 24.
FT DOMAIN 968 1004 EGF-LIKE 25.
FT DOMAIN 1006 1044 EGF-LIKE 26.
FT DOMAIN 1046 1085 EGF-LIKE 27.
FT DOMAIN 1087 1126 EGF-LIKE 28.
FT DOMAIN 1130 1171 EGF-LIKE 29.
FT DOMAIN 1472 1476 POLY-ARG.
FT REPEAT 1165 1212 LTN/NOTCH 1.
FT REPEAT 1233 1246 LTN/NOTCH 2.
FT REPEAT 1247 1286 LTN/NOTCH 3.
FT REPEAT 1633 1665 ANK 1.
FT REPEAT 1666 1698 ANK 2.
FT REPEAT 1700 1732 ANK 3.
FT REPEAT 1733 1765 ANK 4.
FT REPEAT 1766 1798 ANK 5.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 68 80 BY SIMILARITY.
FT DISULFID 74 103 BY SIMILARITY.
FT DISULFID 105 114 BY SIMILARITY.
FT DISULFID 122 133 BY SIMILARITY.
FT DISULFID 127 143 BY SIMILARITY.
FT DISULFID 145 154 BY SIMILARITY.
FT DISULFID 160 171 BY SIMILARITY.
FT DISULFID 180 187 BY SIMILARITY.
FT DISULFID 182 191 BY SIMILARITY.
FT DISULFID 198 211 BY SIMILARITY.

Query Match 3.1%; Score 151.5; DB 1; Length 2003;
Best Local Similarity 27.9%; Pred. No. 0.36;
Matches 39; Conservative 12; Mismatches 30; Indels 59; Gaps 7;

QY 434 PVPNC---AEGCPGSMIKGYCDPAACNNASCDMDGDCSGSGSRYIAGGGTGSIGV 489
Db 1166 PGPRCCKPKAKGCEGR-SCGDGACDAGCSGPGGNNMGDC-----SIGV 1207
QY 490 GHPWQGGGINSVSYC-----NOGCA-----NSWLAKYF- 518
Db 1208 PDPMWK---GCPSHSRWMLFRDQCHPCQCSSECELFQGYDCRTPACTPAYDQYCHDFH 1264
QY 519 ----CDQACNVLSGSPDAGDC 535
Db 1265 NGHCCKGKNTAEQGMGSDC 1284

RESULT 11
NOTC_XENLA STANDARD; PRT; 2524 AA.
ID NOTC_XENLA
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (NOTCH protein).
GN NOTCH.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCB1_TaxID=8355;
RN NCB1
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Notch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.
CC -1- SIMILARITY: Belongs to the NOTCH family.
```

FT	DOMAIN	1020	1056	EGF-LIKE 27,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1058	1094	EGF-LIKE 28,	
FT	DOMAIN	1096	1142	EGF-LIKE 29,	
FT	DOMAIN	1144	1180	EGF-LIKE 30,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1182	1218	EGF-LIKE 31,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1220	1264	EGF-LIKE 32,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1266	1304	EGF-LIKE 33,	
FT	DOMAIN	1306	1346	EGF-LIKE 34,	
FT	DOMAIN	1347	1383	EGF-LIKE 35,	
FT	DOMAIN	1386	1424	EGF-LIKE 36,	
FT	REPEAT	1441	1478	LIN/NOTCH 1,	
FT	REPEAT	1479	1520	LIN/NOTCH 2,	
FT	REPEAT	1521	1560	LIN/NOTCH 3,	
FT	REPEAT	1876	1919	ANK 1,	
FT	REPEAT	1924	1953	ANK 2,	
FT	REPEAT	1957	1987	ANK 3,	
FT	REPEAT	1991	2020	ANK 4,	
FT	REPEAT	2024	2053	ANK 5,	
FT	REPEAT	2057	2086	ANK 6,	
FT	DISULFID	22	35	BY SIMILARITY,	
FT	DISULFID	29	45	BY SIMILARITY,	
FT	DISULFID	47	56	BY SIMILARITY,	
FT	DISULFID	62	74	BY SIMILARITY,	
FT	DISULFID	68	87	BY SIMILARITY,	
FT	DISULFID	89	98	BY SIMILARITY,	
FT	DISULFID	106	117	BY SIMILARITY,	
FT	DISULFID	111	128	BY SIMILARITY,	
FT	DISULFID	130	139	BY SIMILARITY,	
FT	DISULFID	145	156	BY SIMILARITY,	
FT	DISULFID	150	165	BY SIMILARITY,	
FT	DISULFID	167	176	BY SIMILARITY,	
FT	DISULFID	183	194	BY SIMILARITY,	
FT	DISULFID	188	203	BY SIMILARITY,	
FT	DISULFID	205	214	BY SIMILARITY,	
FT	DISULFID	221	232	BY SIMILARITY,	
FT	DISULFID	226	242	BY SIMILARITY,	
FT	DISULFID	244	253	BY SIMILARITY,	
FT	DISULFID	260	271	BY SIMILARITY,	
FT	DISULFID	265	280	BY SIMILARITY,	
FT	DISULFID	282	291	BY SIMILARITY,	
FT	DISULFID	298	311	BY SIMILARITY,	
FT	DISULFID	305	320	BY SIMILARITY,	
FT	DISULFID	322	331	BY SIMILARITY,	
FT	DISULFID	338	349	BY SIMILARITY,	
FT	DISULFID	343	358	BY SIMILARITY,	
FT	DISULFID	360	369	BY SIMILARITY,	
FT	DISULFID	375	386	BY SIMILARITY,	
FT	DISULFID	380	397	BY SIMILARITY,	
FT	DISULFID	399	408	BY SIMILARITY,	
FT	DISULFID	415	428	BY SIMILARITY,	
FT	DISULFID	422	437	BY SIMILARITY,	
FT	DISULFID	439	448	BY SIMILARITY,	
FT	DISULFID	455	466	BY SIMILARITY,	
FT	DISULFID	460	475	BY SIMILARITY,	
FT	DISULFID	477	486	BY SIMILARITY,	
FT	DISULFID	493	504	BY SIMILARITY,	
FT	DISULFID	498	513	BY SIMILARITY,	
FT	DISULFID	515	524	BY SIMILARITY,	
FT	DISULFID	531	542	BY SIMILARITY,	
FT	DISULFID	536	551	BY SIMILARITY,	
FT	DISULFID	553	562	BY SIMILARITY,	
FT	DISULFID	569	579	BY SIMILARITY,	
FT	DISULFID	574	588	BY SIMILARITY,	
FT	DISULFID	590	599	BY SIMILARITY,	
FT	DISULFID	606	617	BY SIMILARITY,	
FT	DISULFID	611	626	BY SIMILARITY,	
FT	DISULFID	628	637	BY SIMILARITY,	
FT	DISULFID	644	654	BY SIMILARITY,	
FT	DISULFID	649	663	BY SIMILARITY,	
FT	DISULFID	665	674	BY SIMILARITY,	
FT	DISULFID	681	692	BY SIMILARITY,	
FT	DISULFID	686	701	BY SIMILARITY,	
FT	DISULFID	703	712	BY SIMILARITY,	

FT	DISULFID	719	729	BY SIMILARITY,	
FT	DISULFID	724	738	BY SIMILARITY,	
FT	DISULFID	740	748	BY SIMILARITY,	
FT	DISULFID	756	767	BY SIMILARITY,	
FT	DISULFID	761	776	BY SIMILARITY,	
FT	DISULFID	778	787	BY SIMILARITY,	
FT	DISULFID	794	805	BY SIMILARITY,	
FT	DISULFID	799	814	BY SIMILARITY,	
FT	DISULFID	816	825	BY SIMILARITY,	
FT	DISULFID	832	843	BY SIMILARITY,	
FT	DISULFID	837	854	BY SIMILARITY,	
FT	DISULFID	856	865	BY SIMILARITY,	
FT	DISULFID	872	883	BY SIMILARITY,	
FT	DISULFID	877	892	BY SIMILARITY,	
FT	DISULFID	894	903	BY SIMILARITY,	
FT	DISULFID	910	921	BY SIMILARITY,	
FT	DISULFID	915	930	BY SIMILARITY,	
FT	DISULFID	932	941	BY SIMILARITY,	
FT	DISULFID	966	977	BY SIMILARITY,	
FT	DISULFID	991	1006	BY SIMILARITY,	
FT	DISULFID	1008	1017	BY SIMILARITY,	
FT	DISULFID	1024	1035	BY SIMILARITY,	
FT	DISULFID	1029	1044	BY SIMILARITY,	
FT	DISULFID	1046	1055	BY SIMILARITY,	
FT	DISULFID	1062	1073	BY SIMILARITY,	
FT	DISULFID	1067	1082	BY SIMILARITY,	
FT	DISULFID	1084	1093	BY SIMILARITY,	
FT	DISULFID	1100	1121	BY SIMILARITY,	
FT	DISULFID	1115	1130	BY SIMILARITY,	

Query Match 3.0%; Score 149.5; DB 1; Length 2524;
Best Local Similarity 19.6%; Pred. No. 0.67; Mismatches 217; Indels 275; Gaps 42;
Matches 145; Conservative 95;

Qy	412	GRDWPDPDFSHSKGVYLTWVPNCABGCGSWIKDS-YCDKACNNACDMDGDCSG	470
Db	1435	GKNIPTDNDICENQ-----CSE-----LADKVCNACNNACGMDGDC--	1477
Qy	471	NSGGSRYIAGCGTSGVGHPEWQ-----FGGINSVSYCNC-GCA-----	510
Db	1478	-----SLNPNDFWKNCTGSLQCKRYFNDD-KDSCQNNNGCYDGFDCQK	1521
Qy	511	-----NSWLADKF-----CDQACNVISCGFDAGDCQ--DHFE--LYVILLP--	550
Db	1522	VEVQCNPVLDQYCKHFGDGHCDQGCNNACCEMDGLDCAENMENAEGTLVLVLMPPR	1581
Qy	551	---NQHVI-----IPKE--CLPYF-SFAEYAKGVGGA---YSDNP--	584
Db	1582	LKNNVNFRELRLVLTWVVFVKDSKGEYKIPYVYGNEEELKHKIKRSTWSDAPSA	1641
Qy	585	-----IIRHASTANK--WKTILHMSGMNATTIHFNLTJF	617
Db	1642	IFSTKESILLGHRRELDMEVRGSIYLEIDNRQCYSSQCFNSATDVAFGLALAS	1701
Qy	618	QNTNDE-EFMQITVEVDTRREGPKLNSTAKGYENLVSEITLL-----PEA	662
Db	1702	LGSLDLTLYKIE-AVXSENNMETPK--PSTLYPMLMLVPLILIFFMVVIYVKRRREHG	1759
Qy	663	EILFED--IPKEKRFKFKRHD-----VNST	686
Db	1760	QLMFPGCFIPKEP--SKKRRRDLGSDYGLKPIKNMTDGSFMDNQNQEWGDEETLENKR	1817
Qy	687	RRQAEVKIP-LVINSLIFKD-----AQLSLNTL-----DLQLEHGDITLKG--	727
Db	1818	FRFEQVILPELVDDTDTRQMTROGLDAADRISMATPPQGEIADCMVNVRRGPG	1877
Qy	728	-----YLSKSALEBSFLANSQAKIKNOAILITDENDSLVAPQ	766
Db	1878	FPLMLASCGGGLETGNSBEEEDSANNISDPI--GQGAQLAHQTDRTGTALHLAAY	1935
Qy	767	EK-QVHKSILPNS--LGVERLQRLTPPVSVKVNGHDG-----QNPPLDET-----	812
Db	1936	ARADAKRLLESADANVQDMNGRTPLHA---VAADAGVQIILIRNATLDDLRMPDG	1992

```
QY 813 -----TARFVETHTQKTTGG-----NTKKEKPS-----LIYPLBSQMTKEKKTGKE 856
Dy 1993 TTPLLIARLAVEGVNEELINAHADVNAVDEFGKSAHMAAVNNVDAAVLLKNSANKD 2052
QY 857 KENSHMEEN-----AENHIGVTEVLLGRKLOHYTD-----SYLGLPWE--KKKTFOD--- 902
Dy 2053 MNNNEETSLFLAAEESYETAKVL---LDHYNRDITDMDLPRDIAGERHNIIVH 2108
QY 903 LIDEEESLKT 912
Dy 2109 LIDEEYNLVKS 2118

RESULT 12
NTC3_RAT
ID NTC3_RAT STANDARD; PRT; 2319 AA.
AC Q9RI72;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN NOTCH3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann T., Boulter J., Weinmaster G., Schanen N.C.;
RT "Rattus norvegicus mRNA for Notch 3."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP FUNCTION.
RA MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Honjo T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT neural progenitor cells to an astroglial fate."
RL Neuron 29:45-55(2001).
[3]
RP TISSUE SPECIFICITY.
RX MEDLINE=2131789; PubMed=11438922;
RX Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RA "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT functional roles for the Notch-DSL signaling system during brain
RT development."
RL J. Comp. Neurol. 436:167-181(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC tagged, jagged2 and Deltal to regulate cell-fate determination.
CC Upon ligand activation through the released core intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
CC (CNS) germinal zones and, in early postnatal life, within
CC numerous cells throughout the CNS. It is more highly localized
CC to ventricular germinal zones.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
```

```
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 34 EGF-like domains.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF164486; AAD46653.2; -.
DR HSSP; P00740; IEDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxy1_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_III.
DR InterPro; IPR002049; LamInin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 33.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF02279; Notch; 1.
DR PRINTS; PRO0010; EGFBL00D.
DR PRINTS; PRO0011; EGFFLAMININ.
DR PRINTS; PRO1452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_Ca; 20.
DR PROSITE; PS50297; ANK_REPEAT; 4.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 18.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS50026; EGF_3; 34.
DR PROSITE; PS01187; EGF_Ca; 16.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 40
FT CHAIN 1 2319
FT CHAIN 1631 2319
FT CHAIN 1664 2319
FT DOMAIN 41 1645
FT TRANSMEM 1646 1666
FT DOMAIN 1667 2319
FT DOMAIN 41 79
FT DOMAIN 80 120
FT DOMAIN 121 158
FT DOMAIN 160 197
FT DOMAIN 199 236
FT DOMAIN 238 274
FT DOMAIN 276 314
FT DOMAIN 316 352
FT DOMAIN 353 391
FT DOMAIN 393 431
FT DOMAIN 433 469
FT DOMAIN 471 507
FT DOMAIN 509 545
FT DOMAIN 547 582
```

FT	DOMAIN	584	620	EGF-LIKE 15.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	622	657	EGF-LIKE 16.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	659	695	EGF-LIKE 17.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	697	732	EGF-LIKE 18.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	736	772	EGF-LIKE 19.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	773	810	EGF-LIKE 20.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	812	849	EGF-LIKE 21.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	851	887	EGF-LIKE 22.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	889	924	EGF-LIKE 23.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	926	962	EGF-LIKE 24.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	964	1000	EGF-LIKE 25.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1002	1036	EGF-LIKE 26.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1038	1084	EGF-LIKE 27.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1086	1122	EGF-LIKE 28.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1124	1160	EGF-LIKE 29.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1162	1205	EGF-LIKE 30.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1207	1246	EGF-LIKE 31.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1248	1289	EGF-LIKE 32.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1291	1327	EGF-LIKE 33.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1337	1375	EGF-LIKE 34.	CALCIUM-BINDING (POTENTIAL).
FT	REPEAT	1384	1420	LIN/NOTCH 1.	CALCIUM-BINDING (POTENTIAL).
FT	REPEAT	1425	1461	LIN/NOTCH 2.	CALCIUM-BINDING (POTENTIAL).
FT	REPEAT	1469	1503	LIN/NOTCH 3.	CALCIUM-BINDING (POTENTIAL).
FT	REPEAT	1540	1569	ANK 1.	CALCIUM-BINDING (POTENTIAL).
FT	REPEAT	1573	1593	ANK 2.	CALCIUM-BINDING (POTENTIAL).
FT	REPEAT	1607	1636	ANK 3.	CALCIUM-BINDING (POTENTIAL).
FT	REPEAT	1699	1735	ANK 4.	CALCIUM-BINDING (POTENTIAL).
FT	REPEAT	1973	2002	ANK 5.	CALCIUM-BINDING (POTENTIAL).
FT	SITE	1573	2002	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY).	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	44	56	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	50	67	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	69	78	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	84	95	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	89	108	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	110	119	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	125	136	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	130	146	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	148	157	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	164	176	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	170	185	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	187	196	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	203	214	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	220	224	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	226	235	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	242	253	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	247	262	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	264	273	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	280	293	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	287	302	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	304	313	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	320	331	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	335	340	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	342	351	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	357	368	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	362	379	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	381	390	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	397	410	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	404	419	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	421	430	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	437	448	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	442	457	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	459	468	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	475	486	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	480	495	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	497	506	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	513	524	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	518	533	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	535	544	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	551	561	BY SIMILARITY.	CALCIUM-BINDING (POTENTIAL).

Query Match: 3.0%; Score 148; DB 1; Length 2319;
 Best Local Similarity: 25.9%; Pred. No. 0.73;

Matches	36;	Conservative	12;	Mismatches	35;	Indels	56;	Gaps	5;
Qy	436	PNCABGCPGSMIDGICDCAACNNACDPMGDCSGSRIYAGGCTGSGVGHFWO-							494
Db	1387	PRCPRAACAKRBDQDCRBNSPGCGWGGDC-							1429
Qy	495	-----GGGINSYSYCN-----QCCASWLDKPC							519
Db	1430	CEALQCMRLFNNSRCBPACSSPACLYDNEPCISGGDR--CNPYKTKCALHF-ADGRG							1486
Qy	520	DQACNVLSGCFDAGCGGD	538						
Db	1487	DQGCNTEGCGWGLDCASE	1505						
RESULT 13									
NTC3_MOUSE	STANDARD;	PRT;	2318	AA.					
ID	Q61982;								
DT	01-NOV-1997 (Rel. 35, Created)								
DT	01-NOV-1997 (Rel. 35, Last sequence update)								
DT	28-FEB-2003 (Rel. 41, Last annotation update)								
DE	Neurogenic locus notch homolog protein 3 precursor (Notch 3).								
GN	NOTCH3.								
OS	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
OX	NCBI_Taxid=10090;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=ICR X Swiss Webster;								
RX	MEDLINE=95001556; PubMed=7918097;								
RA	Lardelli M., Daltstrand J., Lendahl U.;								
RT	"The novel Notch homologue mouse Notch 3 lacks specific epidermal								
RT	growth factor-repeats and is expressed in proliferating								
RT	neuroepithelium.";								
RL	Mech. Dev. 46:123-136(1994).								
RN	[2]								
RP	POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1664.								
RX	MEDLINE=21523956; PubMed=11518718;								
RA	Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;								
RT	"Murine notch homologs (NL-4) undergo presenilin-dependent								
RT	proteolysis.";								
RL	J. Biol. Chem. 276:40268-40273(2001).								
RN	[3]								
RP	POST-TRANSLATIONAL PROCESSING.								
RX	MEDLINE=21374376; PubMed=11459941;								
RA	Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;								
RT	"Conservation of the biochemical mechanisms of signal transduction								
RT	among mammalian Notch family members.";								
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).								
CC	-1- FUNCTION: Functions as a receptor for membrane-bound ligands								
CC	Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.								
CC	Upon ligand activation through the released notch intracellular								
CC	domain (NICD) it forms a transcriptional activator complex with								
CC	RBP-J kappa and activates genes of the enhancer of split locus.								
CC	Affects the implementation of differentiation, proliferation and								
CC	apoptotic programs (By similarity). May play a role during CNS								
CC	development.								
CC	-1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-								
CC	terminal fragment N(IEC) which are probably linked by disulfide								
CC	bonds.								
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein. Following								
CC	proteolytical processing NICD is translocated to the nucleus.								
CC	-1- TISSUE SPECIFICITY: Proliferating neuroepithelium.								
CC	-1- DEVELOPMENTAL STAGE: CNS development.								
CC	-1- PTM: Synthesized in the endoplasmic reticulum as an inactive form								
CC	which is proteolytically cleaved by a furin-like convertase in the								
CC	trans-golgi network before it reaches the plasma membrane to yield								
CC	an active, ligand-accessible form. Cleavage results in a C-								
CC	terminal fragment N(TW) and a N-terminal fragment N(IEC). Following								
CC	ligand binding, it is cleaved by TNF-alpha converting enzyme								
CC	(TACE) to yield a membrane-associated intermediate fragment called								

Best Local Similarity 25.9%; Pred. No. 0.85;
Matches 36; Conservative 11; Mismatches 36; Indels 56; Gaps 5;

QY 436 PNOAGSGPQSWIKDVCDAKNNNSACDMDGPGSGNSGSRRIAGGGTGVCHPMQ- 494
1386 PRCPRACQAKRGSDQNCRENTPGCGMDGDC-----SUNVDDPRQ 1428

QY 495 -----FGGINSVSYCN-----GCGANSMWLADKFC 519
1429 CEALQCMRLFNNSRCDPACSPACLYDNPDGYSGRDRP--CNPYKXCADHF-ADGRC 1485

QY 520 DOACNVLSCGPDAGCGCD 538
1486 DQGCNTEBCGMDGLDCASE 1504

Db

RESULT 14
NTCL HUMAN
ID NTCL HUMAN STANDARD; PRT: 2556 AA.

AC P46531;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (HNI)
DE (Translocation-associated notch protein TAN-1).
GN NOTCH1 OR TAN1.

OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid:9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Mann R.S., Blummeier C.M., Zagouras P.;
RT "Complete human notch 1 (hnt) cDNA sequence."
RT Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-2444 FROM N.A.
RA MEDLINE=91347367; PubMed=1831692;
RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms."
RL Cell 66:649-661(1991).
RN [3]
RP IDENTIFICATION OF LIGANDS.
RA MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.L.,
RT Banks A., Leiman U., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RL "Human ligands of the Notch receptor."
RN [4]
RN INTERACTION WITH DTX1.
RP MEDLINE=98250176; PubMed=9590294;
RA Matsuno K., Eastman D., Mitsiadis T., Quinn A.M., Carcanciu M.L.,
RT Ordenitch P., Kadesch T., Artavanis-Tsakonas S.;
RL "Human deltex is a conserved regulator of Notch signaling."
RN Nat. Genet. 19:74-78(1998).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs. May be important for normal lymphocyte
CC function. In altered form, may contribute to transformation or
CC progression in some T-cell neoplasms. Involved in the maturation
CC of both CD4+ and CD8+ cells in the thymus (By similarity).
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(IEC) which are probably linked by disulfide
CC bonds (By similarity). Interacts with DTX1 and DTX2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus (By

CC similarity).
CC -1- TISSUE SPECIFICITY: In fetal tissues most abundant in spleen,
CC brain stem and lung. Also present in most adult tissues where it
CC is found mainly in lymphoid tissues.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(IEC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXTR). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- DISEASE: NOTCH1 truncation is associated with T-cell acute
CC lymphoblastic leukemia.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 5 ANK repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF308602; AAC33848.1; -.
DR EMBL: M73980; AAA60614.1; -.
DR HSSP: P00740; IEDM.
DR Genew: HGNC:7881; NOTCH1.
DR MIM: 190198; -.
DR GO: GO:0016021; C:integral to membrane; NAS.
DR GO: GO:0003793; P:defense/immunity protein activity; NAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl_S.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR002049; laminin_EGF.
DR InterPro: IPR008287; Notch.
DR InterPro: IPR008000; Notch_dom.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00066; notch; 3.
DR Pfam: PF00008; EGF; 35.
DR PIRSF: PIRSF002279; Notch; 1.
DR PRINTS: PRO0010; EGFAMININ.
DR PRINTS: PRO0011; EGFAMININ.
DR PRINTS: PRO1452; NOTCH.
DR SMART: SM00248; ANK; 6.
DR SMART: SM00179; EGF_CA; 23.
DR SMART: SM00004; NL; 3.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50088; ANK_REPEAT; 4.
DR PROSITE: PS00010; ASK_HYDROXYL; 20.
DR PROSITE: PS00022; ASK_1; 34.
DR PROSITE: PS01186; EGF_2; 26.
DR PROSITE: PS50026; EGF_3; 36.
DR PROSITE: PS01187; EGF_CA; 18.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1..18
FT CHAIN 19..2556 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT CHAIN 1722..2556 NOTCH EXTRACELLULAR TRUNCATION (BY
FT SIMILARITY).
FT CHAIN 1755..2556 NOTCH INTRACELLULAR DOMAIN (BY

```

FT  DOMAIN 19 1736 SIMILARITY)
FT  TRASMEN 1737 1757 EXTRACELLULAR (POTENTIAL).
FT  DOMAIN 1758 2556 POTENTIAL.
FT  DOMAIN 20 58 CYTOPLASMIC (POTENTIAL).
FT  DOMAIN 59 99 EGF-LIKE 1.
FT  DOMAIN 102 139 EGF-LIKE 2.
FT  DOMAIN 120 139 EGF-LIKE 3.
FT  DOMAIN 140 176 EGF-LIKE 4.
FT  DOMAIN 176 216 EGF-LIKE 5.
FT  DOMAIN 218 255 EGF-LIKE 6.
FT  DOMAIN 257 293 EGF-LIKE 7.
FT  DOMAIN 293 333 EGF-LIKE 8.
FT  DOMAIN 335 371 EGF-LIKE 9.
FT  DOMAIN 372 410 EGF-LIKE 10.
FT  DOMAIN 412 450 EGF-LIKE 11.
FT  DOMAIN 452 488 EGF-LIKE 12.
FT  DOMAIN 490 526 EGF-LIKE 13.
FT  DOMAIN 528 564 EGF-LIKE 14.
FT  DOMAIN 566 601 EGF-LIKE 15.
FT  DOMAIN 603 639 EGF-LIKE 16.
FT  DOMAIN 641 676 EGF-LIKE 17.
FT  DOMAIN 678 714 EGF-LIKE 18.
FT  DOMAIN 716 751 EGF-LIKE 19.
FT  DOMAIN 753 789 EGF-LIKE 20.
FT  DOMAIN 791 827 EGF-LIKE 21.
FT  DOMAIN 829 868 EGF-LIKE 22.
FT  DOMAIN 870 906 EGF-LIKE 23.
FT  DOMAIN 908 944 EGF-LIKE 24.
FT  DOMAIN 946 982 EGF-LIKE 25.
FT  DOMAIN 984 1020 EGF-LIKE 26.
FT  DOMAIN 1022 1058 EGF-LIKE 27.
FT  DOMAIN 1060 1096 EGF-LIKE 28.
FT  DOMAIN 1098 1144 EGF-LIKE 29.
FT  DOMAIN 1146 1182 EGF-LIKE 30.
FT  DOMAIN 1184 1220 EGF-LIKE 31.
FT  DOMAIN 1222 1266 EGF-LIKE 32.
FT  DOMAIN 1268 1306 EGF-LIKE 33.
FT  DOMAIN 1308 1347 EGF-LIKE 34.
FT  DOMAIN 1349 1385 EGF-LIKE 35.
FT  DOMAIN 1388 1427 EGF-LIKE 36.
FT  REPEAT 1446 1481 LIN/NOTCH 1.
FT  REPEAT 1482 1523 LIN/NOTCH 2.
FT  REPEAT 1524 1563 LIN/NOTCH 3.
FT  REPEAT 1564 1597 ANK 1.
FT  REPEAT 1598 1631 ANK 2.
FT  REPEAT 1632 1665 ANK 3.
FT  REPEAT 1666 1699 ANK 4.
FT  REPEAT 1700 1733 ANK 5.
FT  REPEAT 1734 1767 POLY-VAL.
FT  REPEAT 1768 1801 POLY-ARG.
FT  REPEAT 1802 1835 POLY-PRO.
FT  REPEAT 1836 1869 POLY-ALA.
FT  REPEAT 1870 1903 POLY-GLU.
FT  REPEAT 1904 1937 POLY-GLY.
FT  REPEAT 1938 1971 POLY-GLN.
FT  REPEAT 1972 2005 POLY-PRO.
FT  REPEAT 2006 2039 POLY-SER.
FT  REPEAT 2040 2073 CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY
FT  REPEAT 2074 2107 SIMILARITY).
FT  REPEAT 2108 2141 BY SIMILARITY.
FT  REPEAT 2142 2175 BY SIMILARITY.
FT  REPEAT 2176 2209 BY SIMILARITY.
FT  REPEAT 2210 2243 BY SIMILARITY.
FT  REPEAT 2244 2277 BY SIMILARITY.
FT  REPEAT 2278 2311 BY SIMILARITY.
FT  REPEAT 2312 2345 BY SIMILARITY.
FT  REPEAT 2346 2379 BY SIMILARITY.
FT  REPEAT 2380 2413 BY SIMILARITY.
FT  REPEAT 2414 2447 BY SIMILARITY.
FT  REPEAT 2448 2481 BY SIMILARITY.
FT  REPEAT 2482 2515 BY SIMILARITY.
FT  REPEAT 2516 2549 BY SIMILARITY.
FT  REPEAT 2550 2583 BY SIMILARITY.
FT  REPEAT 2584 2617 BY SIMILARITY.
FT  REPEAT 2618 2651 BY SIMILARITY.
FT  REPEAT 2652 2685 BY SIMILARITY.
FT  REPEAT 2686 2719 BY SIMILARITY.
FT  REPEAT 2720 2753 BY SIMILARITY.
FT  REPEAT 2754 2787 BY SIMILARITY.
FT  REPEAT 2788 2821 BY SIMILARITY.
FT  REPEAT 2822 2855 BY SIMILARITY.
FT  REPEAT 2856 2889 BY SIMILARITY.
FT  REPEAT 2890 2923 BY SIMILARITY.
FT  REPEAT 2924 2957 BY SIMILARITY.
FT  REPEAT 2958 2991 BY SIMILARITY.
FT  REPEAT 2992 3025 BY SIMILARITY.
FT  REPEAT 3026 3059 BY SIMILARITY.
FT  REPEAT 3060 3093 BY SIMILARITY.
FT  REPEAT 3094 3127 BY SIMILARITY.
FT  REPEAT 3128 3161 BY SIMILARITY.
FT  REPEAT 3162 3195 BY SIMILARITY.
FT  REPEAT 3196 3229 BY SIMILARITY.
FT  REPEAT 3230 3263 BY SIMILARITY.
FT  REPEAT 3264 3297 BY SIMILARITY.
FT  REPEAT 3298 3331 BY SIMILARITY.
FT  REPEAT 3332 3365 BY SIMILARITY.
FT  REPEAT 3366 3399 BY SIMILARITY.
FT  REPEAT 3400 3433 BY SIMILARITY.
FT  REPEAT 3434 3467 BY SIMILARITY.
FT  REPEAT 3468 3501 BY SIMILARITY.
FT  REPEAT 3502 3535 BY SIMILARITY.
FT  REPEAT 3536 3569 BY SIMILARITY.
FT  REPEAT 3570 3603 BY SIMILARITY.
FT  REPEAT 3604 3637 BY SIMILARITY.
FT  REPEAT 3638 3671 BY SIMILARITY.
FT  REPEAT 3672 3705 BY SIMILARITY.
FT  REPEAT 3706 3739 BY SIMILARITY.
FT  REPEAT 3740 3773 BY SIMILARITY.
FT  REPEAT 3774 3807 BY SIMILARITY.
FT  REPEAT 3808 3841 BY SIMILARITY.
FT  REPEAT 3842 3875 BY SIMILARITY.
FT  REPEAT 3876 3909 BY SIMILARITY.
FT  REPEAT 3910 3943 BY SIMILARITY.
FT  REPEAT 3944 3977 BY SIMILARITY.
FT  REPEAT 3978 4011 BY SIMILARITY.
FT  REPEAT 4012 4045 BY SIMILARITY.
FT  REPEAT 4046 4079 BY SIMILARITY.
FT  REPEAT 4080 4113 BY SIMILARITY.
FT  REPEAT 4114 4147 BY SIMILARITY.
FT  REPEAT 4148 4181 BY SIMILARITY.
FT  REPEAT 4182 4215 BY SIMILARITY.
FT  REPEAT 4216 4249 BY SIMILARITY.
FT  REPEAT 4250 4283 BY SIMILARITY.
FT  REPEAT 4284 4317 BY SIMILARITY.
FT  REPEAT 4318 4351 BY SIMILARITY.
FT  REPEAT 4352 4385 BY SIMILARITY.
FT  REPEAT 4386 4419 BY SIMILARITY.
FT  REPEAT 4420 4453 BY SIMILARITY.
FT  REPEAT 4454 4487 BY SIMILARITY.
FT  REPEAT 4488 4521 BY SIMILARITY.
FT  REPEAT 4522 4555 BY SIMILARITY.
FT  REPEAT 4556 4589 BY SIMILARITY.
FT  REPEAT 4590 4623 BY SIMILARITY.
FT  REPEAT 4624 4657 BY SIMILARITY.
FT  REPEAT 4658 4691 BY SIMILARITY.
FT  REPEAT 4692 4725 BY SIMILARITY.
FT  REPEAT 4726 4759 BY SIMILARITY.
FT  REPEAT 4760 4793 BY SIMILARITY.
FT  REPEAT 4794 4827 BY SIMILARITY.
FT  REPEAT 4828 4861 BY SIMILARITY.
FT  REPEAT 4862 4895 BY SIMILARITY.
FT  REPEAT 4896 4929 BY SIMILARITY.
FT  REPEAT 4930 4963 BY SIMILARITY.
FT  REPEAT 4964 4997 BY SIMILARITY.
FT  REPEAT 4998 5031 BY SIMILARITY.
FT  REPEAT 5032 5065 BY SIMILARITY.
FT  REPEAT 5066 5099 BY SIMILARITY.
FT  REPEAT 5100 5133 BY SIMILARITY.
FT  REPEAT 5134 5167 BY SIMILARITY.
FT  REPEAT 5168 5201 BY SIMILARITY.
FT  REPEAT 5202 5235 BY SIMILARITY.
FT  REPEAT 5236 5269 BY SIMILARITY.
FT  REPEAT 5270 5303 BY SIMILARITY.
FT  REPEAT 5304 5337 BY SIMILARITY.
FT  REPEAT 5338 5371 BY SIMILARITY.
FT  REPEAT 5372 5405 BY SIMILARITY.
FT  REPEAT 5406 5439 BY SIMILARITY.
FT  REPEAT 5440 5473 BY SIMILARITY.
FT  REPEAT 5474 5507 BY SIMILARITY.
FT  REPEAT 5508 5541 BY SIMILARITY.
FT  REPEAT 5542 5575 BY SIMILARITY.
FT  REPEAT 5576 5609 BY SIMILARITY.
FT  REPEAT 5610 5643 BY SIMILARITY.
FT  REPEAT 5644 5677 BY SIMILARITY.
FT  REPEAT 5678 5711 BY SIMILARITY.
FT  REPEAT 5712 5745 BY SIMILARITY.
FT  REPEAT 5746 5779 BY SIMILARITY.
FT  REPEAT 5780 5813 BY SIMILARITY.
FT  REPEAT 5814 5847 BY SIMILARITY.
FT  REPEAT 5848 5881 BY SIMILARITY.
FT  REPEAT 5882 5915 BY SIMILARITY.
FT  REPEAT 5916 5949 BY SIMILARITY.
FT  REPEAT 5950 5983 BY SIMILARITY.
FT  REPEAT 5984 6017 BY SIMILARITY.
FT  REPEAT 6018 6051 BY SIMILARITY.
FT  REPEAT 6052 6085 BY SIMILARITY.
FT  REPEAT 6086 6119 BY SIMILARITY.
FT  REPEAT 6120 6153 BY SIMILARITY.
FT  REPEAT 6154 6187 BY SIMILARITY.
FT  REPEAT 6188 6221 BY SIMILARITY.
FT  REPEAT 6222 6255 BY SIMILARITY.
FT  REPEAT 6256 6289 BY SIMILARITY.
FT  REPEAT 6290 6323 BY SIMILARITY.
FT  REPEAT 6324 6357 BY SIMILARITY.
FT  REPEAT 6358 6391 BY SIMILARITY.
FT  REPEAT 6392 6425 BY SIMILARITY.
FT  REPEAT 6426 6459 BY SIMILARITY.
FT  REPEAT 6460 6493 BY SIMILARITY.
FT  REPEAT 6494 6527 BY SIMILARITY.
FT  REPEAT 6528 6561 BY SIMILARITY.
FT  REPEAT 6562 6595 BY SIMILARITY.
FT  REPEAT 6596 6629 BY SIMILARITY.
FT  REPEAT 6630 6663 BY SIMILARITY.
FT  REPEAT 6664 6697 BY SIMILARITY.
FT  REPEAT 6698 6731 BY SIMILARITY.
FT  REPEAT 6732 6765 BY SIMILARITY.
FT  REPEAT 6766 6799 BY SIMILARITY.
FT  REPEAT 6800 6833 BY SIMILARITY.
FT  REPEAT 6834 6867 BY SIMILARITY.
FT  REPEAT 6868 6901 BY SIMILARITY.
FT  REPEAT 6902 6935 BY SIMILARITY.
FT  REPEAT 6936 6969 BY SIMILARITY.
FT  REPEAT 6970 7003 BY SIMILARITY.
FT  REPEAT 7004 7037 BY SIMILARITY.
FT  REPEAT 7038 7071 BY SIMILARITY.
FT  REPEAT 7072 7105 BY SIMILARITY.
FT  REPEAT 7106 7139 BY SIMILARITY.
FT  REPEAT 7140 7173 BY SIMILARITY.
FT  REPEAT 7174 7207 BY SIMILARITY.
FT  REPEAT 7208 7241 BY SIMILARITY.
FT  REPEAT 7242 7275 BY SIMILARITY.
FT  REPEAT 7276 7309 BY SIMILARITY.
FT  REPEAT 7310 7343 BY SIMILARITY.
FT  REPEAT 7344 7377 BY SIMILARITY.
FT  REPEAT 7378 7411 BY SIMILARITY.
FT  REPEAT 7412 7445 BY SIMILARITY.
FT  REPEAT 7446 7479 BY SIMILARITY.
FT  REPEAT 7480 7513 BY SIMILARITY.
FT  REPEAT 7514 7547 BY SIMILARITY.
FT  REPEAT 7548 7581 BY SIMILARITY.
FT  REPEAT 7582 7615 BY SIMILARITY.
FT  REPEAT 7616 7649 BY SIMILARITY.
FT  REPEAT 7650 7683 BY SIMILARITY.
FT  REPEAT 7684 7717 BY SIMILARITY.
FT  REPEAT 7718 7751 BY SIMILARITY.
FT  REPEAT 7752 7785 BY SIMILARITY.
FT  REPEAT 7786 7819 BY SIMILARITY.
FT  REPEAT 7820 7853 BY SIMILARITY.
FT  REPEAT 7854 7887 BY SIMILARITY.
FT  REPEAT 7888 7921 BY SIMILARITY.
FT  REPEAT 7922 7955 BY SIMILARITY.
FT  REPEAT 7956 7989 BY SIMILARITY.
FT  REPEAT 7990 8023 BY SIMILARITY.
FT  REPEAT 8024 8057 BY SIMILARITY.
FT  REPEAT 8058 8091 BY SIMILARITY.
FT  REPEAT 8092 8125 BY SIMILARITY.
FT  REPEAT 8126 8159 BY SIMILARITY.
FT  REPEAT 8160 8193 BY SIMILARITY.
FT  REPEAT 8194 8227 BY SIMILARITY.
FT  REPEAT 8228 8261 BY SIMILARITY.
FT  REPEAT 8262 8295 BY SIMILARITY.
FT  REPEAT 8296 8329 BY SIMILARITY.
FT  REPEAT 8330 8363 BY SIMILARITY.
FT  REPEAT 8364 8397 BY SIMILARITY.
FT  REPEAT 8398 8431 BY SIMILARITY.
FT  REPEAT 8432 8465 BY SIMILARITY.
FT  REPEAT 8466 8499 BY SIMILARITY.
FT  REPEAT 8500 8533 BY SIMILARITY.
FT  REPEAT 8534 8567 BY SIMILARITY.
FT  REPEAT 8568 8601 BY SIMILARITY.
FT  REPEAT 8602 8635 BY SIMILARITY.
FT  REPEAT 8636 8669 BY SIMILARITY.
FT  REPEAT 8670 8703 BY SIMILARITY.
FT  REPEAT 8704 8737 BY SIMILARITY.
FT  REPEAT 8738 8771 BY SIMILARITY.
FT  REPEAT 8772 8805 BY SIMILARITY.
FT  REPEAT 8806 8839 BY SIMILARITY.
FT  REPEAT 8840 8873 BY SIMILARITY.
FT  REPEAT 8874 8907 BY SIMILARITY.
FT  REPEAT 8908 8941 BY SIMILARITY.
FT  REPEAT 8942 8975 BY SIMILARITY.
FT  REPEAT 8976 9009 BY SIMILARITY.
FT  REPEAT 9010 9043 BY SIMILARITY.
FT  REPEAT 9044 9077 BY SIMILARITY.
FT  REPEAT 9078 9111 BY SIMILARITY.
FT  REPEAT 9112 9145 BY SIMILARITY.
FT  REPEAT 9146 9179 BY SIMILARITY.
FT  REPEAT 9180 9213 BY SIMILARITY.
FT  REPEAT 9214 9247 BY SIMILARITY.
FT  REPEAT 9248 9281 BY SIMILARITY.
FT  REPEAT 9282 9315 BY SIMILARITY.
FT  REPEAT 9316 9349 BY SIMILARITY.
FT  REPEAT 9350 9383 BY SIMILARITY.
FT  REPEAT 9384 9417 BY SIMILARITY.
FT  REPEAT 9418 9451 BY SIMILARITY.
FT  REPEAT 9452 9485 BY SIMILARITY.
FT  REPEAT 9486 9519 BY SIMILARITY.
FT  REPEAT 9520 9553 BY SIMILARITY.
FT  REPEAT 9554 9587 BY SIMILARITY.
FT  REPEAT 9588 9621 BY SIMILARITY.
FT  REPEAT 9622 9655 BY SIMILARITY.
FT  REPEAT 9656 9689 BY SIMILARITY.
FT  REPEAT 9690 9723 BY SIMILARITY.
FT  REPEAT 9724 9757 BY SIMILARITY.
FT  REPEAT 9758 9791 BY SIMILARITY.
FT  REPEAT 9792 9825 BY SIMILARITY.
FT  REPEAT 9826 9859 BY SIMILARITY.
FT  REPEAT 9860 9893 BY SIMILARITY.
FT  REPEAT 9894 9927 BY SIMILARITY.
FT  REPEAT 9928 9961 BY SIMILARITY.
FT  REPEAT 9962 9995 BY SIMILARITY.
FT  REPEAT 9996 10000 BY SIMILARITY.

```

```

Query Match 3.0%; Score 145.5; DB 1; Length 2556;
Best Local Similarity 26.1%; Pred. No. 1.2;
Matches 52; Conservative 15; Mismatches 65; Indels 67; Gaps 11;

Qy 435 VPCAGGCPSPSWTKDQYCDKACNNNSACDPDGGCGSGSGS-----RYIAGG--481
Db 1452 LPEQEDBAGNK-----VCSIQCNHAGCGMDGDCSIFENDPWKNCITQSIQCWRTFSQDHC 1506

Qy 482 ----GGTSGYGVHPWQFGGINSVSYCN-----QGCNSWLADKFCDOACNVLSCGFPDAG 533
Db 1507 DSGCNSAGCLFDFDQDQRAEG-----QCNPVLDQYCKDHF-SDGHCDQGCNSAFCEMDGL 1560

Qy 534 DCGQDFHFE-----LYXITLIP-----NQHYYITPKRECL-----P 564
Db 1561 DCA-EHVPESLHAGTLVAVVLMPEQLRNSFFHLELSRLVLTNVFKRAHGQGMIFP 1619

Qy 565 YFSFAE-----VAKRGVEG 578
Db 1620 YGGBEELRHKPIKRAAEG 1638

RESULT 15
GIDA_BUCBP STANDARD; PRT; 626 AA.
ID GIDA_BUCBP
AC P594B5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Glucose inhibited division protein A.
GN GIDA OR BBP001.
OS Buchnera aphidicola (subsp. Balzonga pistracae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Posrigo M., Silva F.J.,
RA Tamames J., Vignera E., Iatorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -!- FUNCTION: Not known.
CC -!- SIMILARITY: Belongs to the gida family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AB014016; AAC26745.1; -.
DR HAVAB; MF_00128; -.
DR InterPro; IPR002218; GIDA.
DR InterPro; IPR004416; GIDA_sub.
DR Pfam; PF01134; GIDA; 1.
DR TIGRFAMs; TIGR00136; gida; 1.
DR PROSITE; PS01280; GIDA_1; 1.
DR PROSITE; PS01281; GIDA_2; 1.
DR Complete proteome.
SQ SEQUENCE 626 AA; 70393 MW; 9019ADAE8B7C274 CRC64;

Query Match 3.0%; Score 145; DB 1; Length 626;
Best Local Similarity 20.3%; Pred. No. 0.16;
Matches 127; Conservative 90; Mismatches 202; Indels 208; Gaps 31;

Qy 212 LTTDKYVPLGLVLMQDLAFSLGPFPTKRNQKTKLPENLSSKVLQLQYSEASVALLK 271
Db 153 LTTGTFPLGGKIVYIGSKCFLGG-----RINDYSSINLAORLKDLPF-----KIGRLKT 199

Qy 272 NNPKDFQELNKQKRNKNTIDGKELTISPAYLLMDLSAISQSKQDEDDISASRFEDNEELRY 331

```

```

Db      200 GFP-----PRINKAT-----INFSKLD-----VQSDNPLPIF 227
QY      332 SLRSTIERHAPWVWPIFYITNGOIPSMINDNPRTVITVQDVFRLSHLPTESS----- 385
Db      228 STMGNOKEHP-----ROIPTITATNEK-----THEIRKXNLKSPITYSGITGIG 273
QY      386 ---PAIESHIRIBGLSQFIYLANDVWFGKDWPPDDFY-----HS-KG-QK 428
Db      274 PRYCSIEDKIVRFSDBRNAHQIFLEPEGHDIETYPNGISTSLPEDVQVEMTHSIKGLER 333
QY      429 VYLTWVPNCAGCGGSMIKDGYCDKACNNSACDWDGDCSGNSGSRXYIAG---GGGT 484
Db      334 AQIT-----RQYAVEYDYCDPRTLKLTLE-----SKPIEGFLAQGIN 372
QY      485 SSIQVGHFWQFG--GAINSVSYCNOGANSWLADKFCDAQVLSGCFDAGDCGDHFE 542
Db      373 GTTGEBEAAAGLLAGLNASLYASNKC--GWFPNR-----GQ----- 407
QY      543 LYKVILLPNQTHYIIPKGECLPYFSFAEYAKGVEGAYSDNPIRHSIANKKTHILIM 602
Db      408 AYLGVLIID---LCTKGTKEPEYRMTARAEHRLI-LREDNADLRLTNIA--KSMNLD 459
QY      603 HSG-----MNATTIHENLTFONTNDEEFMOITVEVDTRBEGKLN--STAOKGY 649
Db      460 NSRWTRRYVKLSNINKEITRELENLKIRS-----KLVSITBLANPFSIKINTBETAK--- 510
QY      650 ENLVSPTILLPEAEILFEDIPEKRF-PKFKSHDVNSTRAQEEVKIPLVNISLLPKDAQ 708
Db      511 -----DLKRPETINYSTLMLEKKEFPGIK-----DKEAYEQIEIQEKYCGYIKROIK 557
QY      709 LSLNLTLDQLEHGDIITIKGVNYSKASALRSFLMSQHAKIKNOALITDETNDSLVAPQEK 768
Db      558 ATKV-----OLNNDYIVL-----SKIKYKVVVKGLSNEVV---SKL 590
QY      769 QVHKSILPNSLGVSERLQRLTPPAVSV 795
Db      591 NPYK--PYSLGQASRISGITPAISI 614

```

Search completed: July 26, 2004, 11:09:43
 Job time : 19.4568 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:05:03 ; Search time 60.6381 Seconds
(without alignments)
4828.666 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907
Sequence: 1 MFKLLQRQTYTCLSHRYGL.....SLKTQLAYFTDSKNTGRQLK 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_plant: *
10: sp_virus: *
11: sp_virus: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4473	91.2	847	4	O96N13
2	3314	67.5	950	4	O9U1L2
3	837	17.1	490	4	O8ETQ2
4	734.5	15.0	384	11	O6I1340
5	488	9.9	666	5	O8SX14
6	488	9.9	666	5	O9V553
7	360	7.3	71	4	O9HUA5
8	329	6.7	602	16	O9L112
9	324.5	6.6	942	16	O9L114
10	322.5	6.6	586	16	O6S853
11	320.5	6.5	486	2	Q7X4S1
12	314	6.4	541	16	O6S851
13	287	5.8	238	2	O9EVX1
14	287	5.8	238	2	O8GPD3
15	266	5.4	545	2	O84D00
16	266	5.4	545	2	O83U59

17	266	5.4	545	16	O9JWM8	O9JWM8 neisseria m
18	265	5.4	545	2	O68215	O68215 neisseria m
19	260.5	5.3	442	5	O861W6	O861W6 dictyostell
20	259	5.3	545	2	O84C29	O84C29 neisseria m
21	258	5.3	333	2	O8GF72	O8GF72 streptococc
22	258	5.3	483	5	O86RM0	O86RM0 dictyostell
23	256	5.2	532	16	O06628	O06628 mycobacteri
24	256	5.2	532	16	O7U184	O7U184 mycobacteri
25	255.5	5.2	522	16	O50025	O50025 mycobacteri
26	255	5.2	370	2	O84CH1	O84CH1 actinobacil
27	254	5.2	364	2	O8XSB4	O8XSB4 actinobacil
28	252	5.1	397	5	O86HR4	O86HR4 dictyostell
29	249	5.0	337	16	O88XJ7	O88XJ7 lactobacill
30	243	5.0	333	2	O83YR8	O83YR8 streptococc
31	231.5	4.7	364	2	O84ER7	O84ER7 aeromonas h
32	230.5	4.7	366	2	O9RGRO	O9RGRO neisseria m
33	228.5	4.7	373	2	O51I51	O51I51 neisseria m
34	226	4.6	360	2	O84EK9	O84EK9 aeromonas h
35	191	3.9	1476	13	O90285	O90285 carassius a
36	183	3.7	638	13	O42372	O42372 brachydantio
37	175.5	3.6	1455	5	O86FJ9	O86FJ9 caenorhabdi
38	175.5	3.6	2352	5	O61240	O61240 halocynthia
39	175	3.6	3869	5	O86PQ3	O86PQ3 cryptospori
40	173	3.5	2524	5	O9GPA5	O9GPA5 brachyosteo
41	170.5	3.5	762	13	O42373	O42373 brachydantio
42	170.5	3.5	2468	13	O800E4	O800E4 brachydantio
43	168.5	3.4	523	5	O9NE33	O9NE33 leishmania
44	167	3.4	6118	5	O81396	O81396 plasmodium
45	161.5	3.3	2528	13	O8AXP0	O8AXP0 cynops pyrr

ALIGNMENTS

RESULT 1
ID O96N13 PRELIMINARY; PRT; 847 AA.
AC O96N13; 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein FLJ15175 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nimomiyu K., Nagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Futuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsura N., Sato K., Tanikawa M., Yamazaki Y., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto U., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Iwagui T.,
RT "NDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK056137; BAB71102.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; P:cell differentiation; IEA.
DR Pfam; PF00066; notch; 1.
DR SMART; SM00004; NU; 2.
KW Hypothetical protein.
FT NON_TER 847 847
SQ SEQUENCE 847 AA; 95165 MW; 0A115015824733C5 CR64;
CY Query Match 91.2%; Score 4473; DB 4; Length 847;
Best Local Similarity 99.2%; Pred. No. 7.9e-289;
Matches 845; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
1 MFKLLQRQTYTCLSHRYGLVYCVGLGVVTVIYSAFQGEVLMKSRDQHYVLFDSYRNDI 60

```

Db      1 MFELQROVYTCLEHRYGLVYCFGLVVTIVSARFQFGEVLEWRDQYHVLFPDSYRDN1
QY      61 AKSPONRLCLPMPIIDVYTTWVNGTDLLELKLQOVREOMBEQKAMEIIGKXTEPTK 120
Db      61 AKSPONRLCLPMPIIDVYTTWVNGTDLLELKLQOVREOMBEQKAMEIIGKXTEPTK 120
QY      121 KSEKQECLELTHCICIVPMVLDPALPANITLKDVSLYSPSHASDIFNVAKPRNPSTNV 180
Db      121 KSEKQECLELTHCICIVPMVLDPALPANITLKDVSLYSPSHASDIFNVAKPRNPSTNV 180
QY      181 SYVVDSTVDVDAHSGILKSGRQVWNGVLTTEKVEPVLMODLFLSGPPTPEET 240
Db      181 SYVVDSTVDVDAHSGILKSGRQVWNGVLTTEKVEPVLMODLFLSGPPTPEET 240
QY      241 NOLKTLPENLSSKVKLLQLYSEASVALLKLNPPDPOELNKOTKQNTIDGKELTISPA 300
Db      241 NOLKTLPENLSSKVKLLQLYSEASVALLKLNPPDPOELNKOTKQNTIDGKELTISPA 300
QY      301 YLLMPLSAISQKQEDISASRFEDNEELRYSLSRISERHAPWVRNIFVTNGQIPSWLNL 360
Db      301 YLLMPLSAISQKQEDISASRFEDNEELRYSLSRISERHAPWVRNIFVTNGQIPSWLNL 360
QY      361 DNPRTVITVHOVFNLSHLPTFSSPAIESHHRLEGSOKFIYNDVMEGKXVMPDDF 420
Db      361 DNPRTVITVHOVFNLSHLPTFSSPAIESHHRLEGSOKFIYNDVMEGKXVMPDDF 420
QY      421 YHSHSGQKRYLTWPVPCNACGCGSGWIDGYCDKACNNACDMGDCGSGSGSRITAG 480
Db      421 YHSHSGQKRYLTWPVPCNACGCGSGWIDGYCDKACNNACDMGDCGSGSGSRITAG 480
QY      481 GGGTSGIGVHPWQGGGINSYSCNOCGANSWLADKFCDOACNVLSGCFPAGDGGDHF 540
Db      481 GGGTSGIGVHPWQGGGINSYSCNOCGANSWLADKFCDOACNVLSGCFPAGDGGDHF 540
QY      541 HELYKVIILPNQTHYIIPKGECLPYFSPAEVAKRGVEGAYSNDPIIRHASIANKWTIHL 600
Db      541 HELYKVIILPNQTHYIIPKGECLPYFSPAEVAKRGVEGAYSNDPIIRHASIANKWTIHL 600
QY      601 INHSGNATTHFNLTFOUNTDEEFKMQITVEVDREBPKINSTOKXENLVSPITLLP 660
Db      601 INHSGNATTHFNLTFOUNTDEEFKMQITVEVDREBPKINSTOKXENLVSPITLLP 660
QY      661 EAEILFEDIPEKRPFKRHVDNSTRAQEEVKIPLVNISLPPDAQLSINTLIDLOEH 720
Db      661 EAEILFEDIPEKRPFKRHVDNSTRAQEEVKIPLVNISLPPDAQLSINTLIDLOEH 720
QY      721 GDTILKGYLSALSIRFLMNSQAKIKNOAIITDENNDLSVAPQEKVHKSILPNSLG 780
Db      721 GDTILKGYLSALSIRFLMNSQAKIKNOAIITDENNDLSVAPQEKVHKSILPNSLG 780
QY      781 VSERLQRLTFPAVSVKVNHDGQNPDLLETTAFRVEHTOKTIGANTKEKPSLIIV 840
Db      781 VSERLQRLTFPAVSVKVNHDGQNPDLLETTAFRVEHTOKTIGANTKEKPSLIIV 840
QY      841 PLESQMT 847
Db      841 PLESQMT 847

RESULT 2
Q9ULI2 PRELIMINARY; PRT; 950 AA.
AC Q9ULI2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein KIAA1208 (Fragment).
GN KIAA1208
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=20039619; PubMed=10574462;
RA      Nagase T., Ishikawa K., Kikuno R., Hicosawa M., Nomura N., Ohara O.,
RT      "Prediction of the coding sequences of unidentified human genes. XV.
RT      The complete sequences of 100 new cDNA clones from brain which code
RT      for large proteins in vitro.";
RL      DNA Res. 6:337-345(1999).
DR      EMBL; AB030304; BAA86522.2; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005509; F:calcium ion binding; IEA.
DR      GO; GO:0030154; P:cell differentiation; IEA.
DR      InterPro; IPR002048; EF-hand.
DR      InterPro; IPR000800; Notch_dom.
DR      Pfam; PF00036; ethand; 1.
DR      Pfam; PF00066; notch; 1.
DR      SMART; SM00004; NL; 2.
DR      PROSITE; PS00018; EF_HAND; 1.
KM      Hypothetical protein.
FT      NON_TER
SQ      SEQUENCE 950 AA; 108785 MW; F653BDFACEACT7503 CRC64;

Query Match 67.5%; Score 3314; DB 4; Length 950;
Best Local Similarity 99.8%; Pred. No. 1.2e-211;
Matches 621; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      307 SAISQKQEDISASRFEDNEELRYSLSRISERHAPWVRNIFVTNGQIPSWLNLNPRVT 366
Db      1 SAISQKQEDISASRFEDNEELRYSLSRISERHAPWVRNIFVTNGQIPSWLNLNPRVT 60
QY      367 IYTHQVFNLSHLPTFSSPAIESHHRLEGSOKFIYNDVMEGKXVMPDDFYSHSG 426
Db      61 IYTHQVFNLSHLPTFSSPAIESHHRLEGSOKFIYNDVMEGKXVMPDDFYSHSG 120
QY      427 QKRYLTWPVPCNACGCGSGWIDGYCDKACNNACDMGDCGSGSGSRITAGGGGTGS 486
Db      121 QKRYLTWPVPCNACGCGSGWIDGYCDKACNNACDMGDCGSGSGSRITAGGGGTGS 180
QY      487 IGVHPWQGGGINSYSCNOCGANSWLADKFCDOACNVLSGCFPAGDGGDHFELVY 546
Db      181 IGVHPWQGGGINSYSCNOCGANSWLADKFCDOACNVLSGCFPAGDGGDHFELVY 240
QY      547 ILLPNQTHYIIPKGECLPYFSPAEVAKRGVEGAYSNDPIIRHASIANKWTIHLIMHSGM 606
Db      241 ILLPNQTHYIIPKGECLPYFSPAEVAKRGVEGAYSNDPIIRHASIANKWTIHLIMHSGM 300
QY      607 NATTHFNLTFOUNTDEEFKMQITVEVDREBPKINSTOKXENLVSPITLLPEAILP 666
Db      301 NATTHFNLTFOUNTDEEFKMQITVEVDREBPKINSTOKXENLVSPITLLPEAILP 360
QY      667 EDIPEKRPFKRHVDNSTRAQEEVKIPLVNISLPPDAQLSINTLIDLOEHGDTLK 726
Db      361 EDIPEKRPFKRHVDNSTRAQEEVKIPLVNISLPPDAQLSINTLIDLOEHGDTLK 420
QY      727 GYNLSKALSIRFLMNSQAKIKNOAIITDENNDLSVAPQEKVHKSILPNSLVSERLQ 786
Db      421 GYNLSKALSIRFLMNSQAKIKNOAIITDENNDLSVAPQEKVHKSILPNSLVSERLQ 480
QY      787 RLTFPAVSVKVNHDGQNPDLLETTAFRVEHTOKTIGANTKEKPSLIIVLESQM 846
Db      481 RLTFPAVSVKVNHDGQNPDLLETTAFRVEHTOKTIGANTKEKPSLIIVLESQM 540
QY      847 TKEKKTGKREKSRMEENAHNIGTVEVLGRKQIHTYDSYGLPWEKKYFQDLDLDE 906
Db      541 TKEKKTGKREKSRMEENAHNIGTVEVLGRKQIHTYDSYGLPWEKKYFQDLDLDE 600
QY      907 EESLKTQLAYPTDSKNTGRQLK 928
Db      601 EESLKTQLAYPTDSKNTGRQLK 622

```

RESULT 3

```
Q86TQ2
ID Q86TQ2 PRELIMINARY; PRT; 490 AA.
AC Q86TQ2;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to v-naf musclicaponeurotic fibrosarcoma oncogene family,
DE protein B (Avian) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042615; AA042615.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF_HAND; 1.
FT NON TER 1
SQ SEQUENCE 490 AA; 57412 MW; 37B19FE0D1259AD2 CRC64;

Query Match 17.1%; Score 837; DB 4; Length 490;
Best Local Similarity 100.0%; Pred. No. 2.7e-47;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 767 EKQVHKSILPNSLGVSEIRQLTFPAVSVKNGHDQGNPPDLLETTARFEVETHTQKI 826
DB 1 EKQVHKSILPNSLGVSEIRQLTFPAVSVKNGHDQGNPPDLLETTARFEVETHTQKI 60
QY 827 GGNVTKKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIVTEVLIGRKLQHYTD 886
DB 61 GGNVTKKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIVTEVLIGRKLQHYTD 120
QY 887 SYLGFPMWEKKYFQDLDDEESLKTQLAYFTDSKNTGRQLK 928
DB 121 SYLGFPMWEKKYFQDLDDEESLKTQLAYFTDSKNTGRQLK 162

RESULT 4
061340 PRELIMINARY; PRT; 384 AA.
ID 061340;
AC 061340;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Basic domain/leucine zipper transcription factor (Fragment).
GN MAFB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95094266; PubMed=8001130;
RA Cordes S.P., Bash G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor."
RL Cell 79:1025-1034(1994).
DR EMBL; L36434; AAA65688.1; -.
DR PIR; I49528; I49528.
DR MGI; MGI:104555; MafB.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006357; P:regulation of transcription from pol II pro. .; IEA.
DR GO; GO:0007379; P:segment specification; IME.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF_HAND; 1.
FT NON TER 1
SQ SEQUENCE 384 AA; 43754 MW; 51F473C8807A7E55 CRC64;
```

```
Query Match 15.0%; Score 734.5; DB 11; Length 384;
Best Local Similarity 57.9%; Pred. No. 1.2e-40;
Matches 165; Conservative 28; Mismatches 71; Indels 21; Gaps 6;

QY 644 TAQGHENLVSPITLLPEAEILPEDIPEKRFKFRHVDNSTRAQOEYKILPLVNISLL 703
DB 3 TTQKAYESLVSPTPLPQADVPEDEPKERFPKIRRHVDNATNGRFOEBVKIPRNVISLL 62
QY 704 PXAOAISLNTLDQLEHGGDITLKGYNLSKALIRSLFMSQHAKIKNOAIITDETDLSLY 763
DB 63 PKAQVRLSNLDDQLERGDITLKGYNLSKALIRSLFMSQHAKIKNOAIITDETDLSLY 121
QY 764 APOEKQVHKSILPNSLGVSEIRQLTFPAVSVKNGHDQGNPPDLLETTARFEVETHTQ 823
DB 122 VPQENSHRR--PHGFAGEHRSERWTAPATVTVKGRDHALNPPVLETNARL-----AQ 174
QY 824 KTIGAVNTEKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIVTEVLIGRKLQHYTD 883
DB 175 PTLGVTSKENSPLLVPPESHLP-----KEESDRABGNA--VPYKELPGRRCR 224
QY 884 YTDSYLGFPMWEKKYFQDLDDEESLKTQLAYFTDSKNTGRQLK 928
DB 225 IIDA--FCPGKKKKXFQDLDDEESLKTQLAYFTDSKNTGRQLK 266

RESULT 5
08SX14 PRELIMINARY; PRT; 666 AA.
ID 08SX14;
AC 08SX14;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE RE35033P.
GN CG8027.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Drensek D., Farfan D., Fries E.,
RA George R., Gonzalez M., Guatin H., Krenmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuno J., Pacled J., Patgas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceinkler S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089618; AAL90356.1; -.
DR FLYBase; FBgn0033392; CG8027.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00066; notch; 1.
SQ SEQUENCE 666 AA; 77745 MW; F6FDB6D1C1C39248 CRC64;

Query Match 9.9%; Score 488; DB 5; Length 666;
Best Local Similarity 54.2%; Pred. No. 7.1e-24;
Matches 84; Conservative 30; Mismatches 41; Indels 0; Gaps 0;

QY 315 DEDISARPEDENEELRYSRSIERHAPVYANFIVTNGOIPSNLNDNPRVITVHQDVF 374
DB 86 DDKYDPRFDDKVELYSLRSLEKHAAMIRHYIVTNGOIPSNLDSYRVTIVPREVLA 145
QY 375 RINSHLPTFSSPAIEBHIRIBELSGKFIYLVNDVWFGADVWPDPPYSHSKQGYVLTLP 434
DB 146 PDDQQLPTFSSSAIEFLHRIPLSLRFLYLVNDVIFLGAFLYEDLYTBAEGVRYQAWM 205
QY 435 VPNCAGCPGSGWIKDGYCDKACNNSACDWDGDCS 469
DB 206 VPGCALDCPWTYITGDAACDRHCNIDACQFDGDCS 240
```

```

RESULT 6
O9V553 PRELIMINARY; PRT; 666 AA.
ID O9V553
AC O9V553
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG8027-PA.
DE CG8027-PA.
GN CG8027.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champs M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,
RA Abrell J.F., Agapayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck U., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibeagwam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Kechum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spralling A.C., Stapleton M., Strong R., Sun E.,
RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodgett, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celisner S.E.,
RA Clump M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.D.,
RA Harris N.L., Kromtiller B., Marshall B., Milburn G.H., Richter J.,
RA Russo S., Searle S.M.J., Smith E., Snu S., Smutnick F.,
RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA FlyBase;
RT Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

```

```

RN [4]
RP SEQUENCE FROM N.A.
RA FlyBase;
RT Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL: AEO03834; AAF58967.2; -.
DR GO: AEO03834; C:membrane; IEA.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0030154; P:cell differentiation; IEA.
DR InterPro: IPR000800; Notch_dom.
DR Pfam: PF00066; notch; 1.
DR SMART: SM00004; Nt; 1.
SQ SEQUENCE 666 AA; 77731 MW; EA232EC57C754FF6 CRC64;

Query Match 9.9%; Score 488; DB 5; Length 666;
Best Local Similarity 54.2%; Pred. No. 7, 1e-24;
Matches 84; Conservative 30; Mismatches 41; Indels 0; Gaps 0;

QY 315 DEDISAREPDENEELRYSLSRISERHAPWVNFIVTNGQIPSWMLNDPRVTIVHQDYF 374
DB 86 DDXYDPSRFDKKNELNYSLSRLEKKAAMIRHYIVTNGQIPSWMLDYSRVTIVPHEVLA 145
QY 375 RNLSHPITSSSPAIESHIRIEGLSGKFTYLNDDVWFGKDWVDPDEFYSHSKQKTYLTP 434
DB 146 PDDQPLPTSSSAIEFHLRIPPLSRFLYLNDDIFLGALYPEDUYTEAGVRYQAWM 205
QY 435 VPCACGCPGSGWIKDGYCDKACNNSACDWDGDCS 469
DB 206 VPCALDCPWTYIGDGCGRHCHNIDACGPDGDCS 240

RESULT 7
ID O9BUA5 PRELIMINARY; PRT; 71 AA.
AC O9BUA5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TissueSkin;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC002779; AA02779.1; -.
KW Hypothetical protein.
SQ SEQUENCE 71 AA; 8401 MW; D4809F1514A85831 CRC64;

Query Match 7.3%; Score 360; DB 4; Length 71;
Best Local Similarity 100.0%; Pred. No. 8, 3e-17;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFKILOROTYCLSRGYVCEFGVVTIYSAPFGFVYEWGRDOYHVPFDSYRNI 60
DB 1 MFKILOROTYCLSRGYVCEFGVVTIYSAPFGFVYEWGRDOYHVPFDSYRNI 60
QY 61 AGKSFQNR 68
DB 61 AGKSFQNR 68

RESULT 8
ID O9LI12 PRELIMINARY; PRT; 602 AA.
AC O9LI12;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein SC02594.
GN SC02594 OR SCC8.05C.
OS Streptomyces coelicolor.

```



```

QY 29 IDCKETLTPSPALLMDLPAISQSKODEISMSREDEENELRYLSRISIEHAPWAVITV 349
Db 242 VDGND----PAWKOKKQAKGEVYHAESAASJARIISDELRYSIRLSH.FAPMIRNIYV 297
QY 350 TNGQIPSWMLNBPVETVTHQDVEPRLNSHPSPSPASIESSH.RREGLSOKRIYUNDV 409
Db 298 TDQVYAMKREDLPQARIATRHREIFRNPEDLPFTNSISISQLEHHIEGLAHEHLYINDM 357
QY 410 MFKGDVPPDPEYSHSKGQKVTYLTWPVEN-CAEGCPGSMXIDGYCDYACNNS 459

```

```
Db 358 FMGRVAPHSFT-ENGARY--FPSRNRIFOGANAE--TSPVDAACKN 403
RESULT 13
Q9EVX1 PRELIMINARY; PRT; 238 AA.
ID Q9EVX1
AC Q9EVX1
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative hexose transferase.
GN CPSJ.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
NCBI_TaxId=1304;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=20519245; PubMed=11065358;
RA Altmann-Roig E., Mulholland F., Gasson M.J., Griffin A.M.;
RT "The complete cps gene cluster from Streptococcus thermophilus NC8B
RT 2393 involved in the biosynthesis of a new exopolysaccharide.";
RL Microbiology 146:2793-2802(2000).
DR EMBL; Y17900; CAC18360.1; -.
DR GO; GO:0016740; P:transferase activity; IEA.
KW Transferase.
SQ SEQUENCE 238 AA; 29110 MW; A964250B01B78988 CRC64;

Query Match 5.8%; Score 287; DB 2; Length 238;
Best Local Similarity 38.3%; Pred. No. 3.6e-11;
Matches 59; Conservative 27; Mismatches 44; Indels 24; Gaps 3;

Qy 272 NNPKDFOLNKOTKQMTIDGKELTISPAYLMDLSAISQKODEDISASRFEDNEELRY 331
Db 11 NDEPIREKNTKTPNHRKID-----NDED-NVHRYDYCTFNY 47
Qy 332 SIRSIERHAPWVNIFIVTNGQIPSWLNDNPRVTIVTHQDVFNLSHLPTSSPAIESH 391
Db 48 WFRWVERHAPWVNIVLITNGQRPKWLNVNHPKLVKVRHEE-FIPKEYLPIFNASAIEMN 106
Qy 392 IHRIGLSQKFIYLNDDVWFGKDVWPDFFYSK 425
Db 107 IHRIDGISENFVLFNDMTLIDVKYSDFVNEK 140

RESULT 14
Q8GPD3 PRELIMINARY; PRT; 238 AA.
ID Q8GPD3
AC Q8GPD3
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE EPSJ.
GN EPSJ.
OS Streptococcus thermophilus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
NCBI_TaxId=1308;
RN (1)
RP SEQUENCE FROM N.A.
RA Rallu F., Ehrlich D.S., Renault P.;
RT "Diversity of eps operons in Streptococcus thermophilus.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF454496; AAN63705.1; -.
SQ SEQUENCE 238 AA; 29110 MW; A964250B01B78988 CRC64;

Query Match 5.8%; Score 287; DB 2; Length 238;
Best Local Similarity 38.3%; Pred. No. 3.6e-11;
Matches 59; Conservative 27; Mismatches 44; Indels 24; Gaps 3;

Qy 272 NNPKDFOLNKOTKQMTIDGKELTISPAYLMDLSAISQKODEDISASRFEDNEELRY 331
Db 11 NDEPIREKNTKTPNHRKID-----NDED-NVHRYDYCTFNY 47
Qy 332 SIRSIERHAPWVNIFIVTNGQIPSWLNDNPRVTIVTHQDVFNLSHLPTSSPAIESH 391
Db 48 WFRWVERHAPWVNIVLITNGQRPKWLNVNHPKLVKVRHEE-FIPKEYLPIFNASAIEMN 106
Qy 392 IHRIGLSQKFIYLNDDVWFGKDVWPDFFYSK 425
Db 107 IHRIDGISENFVLFNDMTLIDVKYSDFVNEK 140
```

```
Db 11 NDEPIREKNTKTPNHRKID-----NDED-NVHRYDYCTFNY 47
Qy 332 SIRSIERHAPWVNIFIVTNGQIPSWLNDNPRVTIVTHQDVFNLSHLPTSSPAIESH 391
Db 48 WFRWVERHAPWVNIVLITNGQRPKWLNVNHPKLVKVRHEE-FIPKEYLPIFNASAIEMN 106
Qy 392 IHRIGLSQKFIYLNDDVWFGKDVWPDFFYSK 425
Db 107 IHRIDGISENFVLFNDMTLIDVKYSDFVNEK 140

RESULT 15
Q84D00 PRELIMINARY; PRT; 545 AA.
ID Q84D00
AC Q84D00
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE SAGB.
GN SAGB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
NCBI_TaxId=487;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=M4775;
RA Sacchi C.T., Whitney A.A., Mayer L.W., Mothershead E., Popovic T.;
RT "Neisseria meningitidis sacb gene.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY234202; AAC85300.1; -.
SQ SEQUENCE 545 AA; 64134 MW; A3648A3C85A946F CRC64;

Query Match 5.4%; Score 266; DB 2; Length 545;
Best Local Similarity 27.8%; Pred. No. 3.1e-09;
Matches 90; Conservative 53; Mismatches 129; Indels 52; Gaps 13;

Qy 145 LPANITLKDVSLYPSFHSASDIFNVAKPKNPSTNVVYFDSTQKVEDAHSGGLKG-NS 203
Db 97 LPENITLK--PALCILESHKEDFLN-----KFLTISSENLKLYKFNQINPKS 145
Qy 204 KQVWVG-----LTTDKVEPGLV-----LMQDLAFLSGFPPTKETNQLT 245
Db 146 VNEIWTDLFSIAHVDMKLTDRFLSSISQFWRFECKEDKDFI----LPTARKYSR 200
Qy 246 KLPENLSSKVKLQLYSEASVALKLNNPKDFOLNKQTKQMTIDGKELTISPAYLMD 305
Db 201 KLMKH---SIKNNQLFKEG-----IRNYSISSLPYEDDHNDIDLVFTWNSBEDKNQ 251
Qy 306 LSAISQSKD---EDISASRFEDNEELRYSLSIEHAAWVNIFIVTNGQIPSWLNDN 362
Db 252 -ELYKXKPDPSNDATSTRFSLRDELKALRWEMNGSIFIRKIFIVSNCAAPPAMIDLNN 310
Qy 363 PRVIVTHQDVFNLSHLPTSSPAIESHIHREGLSQKFIYLNDDVWFGKDVWPDFFYS 422
Db 311 PKIQWYHERIMQ-SALPTFSHALETSLHHPGISNFIYSNDPFLTKPLKNKNFY 369
Qy 423 HSKGQVYL-TWPEVN--CAEGCP 443
Db 370 SNGIATKRLKLEAWGNVNGECTEGEP 393
```

Search completed: July 26, 2004, 11:11:35
Job time : 63.6381 secs

This Page Blank (uspio)

CC patient suffering from a lysosomal storage disease. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.3e-160;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPANPHMIDRIVMQELQDMPEEEDKTSFHKVRS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPANPHMIDRIVMQELQDMPEEEDKTSFHKVRS 60
QY 61 EDMQAFSFFYYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
DB 61 EDMQAFSFFYYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
QY 121 GLEHMLINSKMLPADITQLNNIPPTQESYVDPNLPVTKSLVTNCKPVTDKIKHAYKDX 180
DB 121 GLEHMLINSKMLPADITQLNNIPPTQESYVDPNLPVTKSLVTNCKPVTDKIKHAYKDX 180
QY 181 NKYRFEIMGEEBIAFGMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHKAQTVKAVL 240
DB 181 NKYRFEIMGEEBIAFGMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHKAQTVKAVL 240
QY 241 RDPFESMFPPIPSQFELPREYRNRFIMHLEQEWRAVYRDKLKEWTCVLAATLIMPTISFF 300
DB 241 RDPFESMFPPIPSQFELPREYRNRFIMHLEQEWRAVYRDKLKEWTCVLAATLIMPTISFF 300
QY 301 AEQIALKKRIFFRRRIHKEASPNRIRV 328
DB 301 AEQIALKKRIFFRRRIHKEASPNRIRV 328

RESULT 2

ADD27813
ID ADD27813 standard; protein; 328 AA.

XX ADD27813;

XX 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase beta subunit.

XX human; protein phosphorylation; soluble GlcNAc-phosphotransferase;

KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.

XX Homo sapiens.

PN US2003119088-A1.

XX 26-JUN-2003.

XX 21-DEC-2001; 2001US-00023888.

XX 21-DEC-2001; 2001US-00023888.

PA (NOVA-) NOVAAZYME PHARM INC.

PI Canfield W, Kudo M;

PI N-PSDB; ADD27811.

DR WPI: 2003-801323/75.

XX Phosphorylating a protein for treating a patient suffering from a

PT lysosomal storage disease e.g. Fabry's disease by contacting the protein

PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated

PT protein.

PS Claim 8; SEQ ID NO 5; 55pp; English.

CC The invention relates to a method of phosphorylating a protein comprising

CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-

CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of the human GlcNAc-phosphotransferase beta subunit.

XX
SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 7; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.3e-160;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPANPHMIDRIVMQELQDMPEEEDKTSFHKVRS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPANPHMIDRIVMQELQDMPEEEDKTSFHKVRS 60
QY 61 EDMQAFSFFYYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
DB 61 EDMQAFSFFYYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
QY 121 GLEHMLINSKMLPADITQLNNIPPTQESYVDPNLPVTKSLVTNCKPVTDKIKHAYKDX 180
DB 121 GLEHMLINSKMLPADITQLNNIPPTQESYVDPNLPVTKSLVTNCKPVTDKIKHAYKDX 180
QY 181 NKYRFEIMGEEBIAFGMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHKAQTVKAVL 240
DB 181 NKYRFEIMGEEBIAFGMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHKAQTVKAVL 240
QY 241 RDPFESMFPPIPSQFELPREYRNRFIMHLEQEWRAVYRDKLKEWTCVLAATLIMPTISFF 300
DB 241 RDPFESMFPPIPSQFELPREYRNRFIMHLEQEWRAVYRDKLKEWTCVLAATLIMPTISFF 300
QY 301 AEQIALKKRIFFRRRIHKEASPNRIRV 328
DB 301 AEQIALKKRIFFRRRIHKEASPNRIRV 328

RESULT 3

ABM01489
ID ABM01489 standard; protein; 328 AA.

XX ABM01489;

XX 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase beta subunit precursor protein.

XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell;

KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;

KW gastrointestinal; human; enzyme; lectin resistant cell;

KW deoxymannojirimycin; kifunensine; glycosylation inhibition.

XX Homo sapiens.

PN US2003124652-A1.

XX 03-JUL-2003.

XX 21-DEC-2001; 2001US-00023889.

XX 21-DEC-2001; 2001US-00023889.

PA (NOVA-) NOVAAZYME PHARM INC.

PI Canfield WM;

PI N-PSDB; AAD62491.

DR WPI: 2003-810984/76.

XX Producing a high mannose glycoprotein for treating lysosomal storage

PT disease, comprises culturing the lectin resistant mammalian cell in the

PT presence of deoxymannojirimycin and kifunensine.

PS Claim 10; Page 18-19; 46pp; English.

XX The invention relates to a method for producing a high mannose
 CC glycoprotein. The method comprises: introducing and expressing a
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;
 CC isolating the cell; culturing the cell in the presence of
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in
 CC gene therapy. The method is useful for producing a high mannose
 CC glycoprotein in a complex carbohydrate deficient cell for treating
 CC lysosomal storage disease. The present sequence is human N-
 CC acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor
 CC protein
 CC
 SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 7; Length 328;
 Best Local Similarity 100.0%; Pred. No. 2,3e-160;
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHPMHMDRIVMQELQDMPPEEFDKTSFHKVRS 60
 DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHPMHMDRIVMQELQDMPPEEFDKTSFHKVRS 60
 QY 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRLTRIHLEPLSLQDLT 120
 DB 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRLTRIHLEPLSLQDLT 120
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPNLPVTKSLVYNCKRVTQKIHKAYKDK 180
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPNLPVTKSLVYNCKRVTQKIHKAYKDK 180
 QY 181 NKYRFEIMGESEELAFKIRTNVSHVVGQLDIDKRNKRFVCINDINDHNHKAQTVKAVL 240
 DB 181 NKYRFEIMGESEELAFKIRTNVSHVVGQLDIDKRNKRFVCINDINDHNHKAQTVKAVL 240
 QY 241 RPFYESMFPPIPSQFELPREYRNFLMHLEQEWRAVRDKLKFWTHCVLATLIMFTIFSEFF 300
 DB 241 RPFYESMFPPIPSQFELPREYRNFLMHLEQEWRAVRDKLKFWTHCVLATLIMFTIFSEFF 300
 QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
 DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328

RESULT 4
 ABW01538
 ID ABW01538 standard; protein; 328 AA.
 AC ABW01538;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human GlcNAc-phosphotransferase beta subunit precursor protein.
 XX
 KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 KW N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PN US2003124653-A1.
 XX
 PD 03-UTL-2003.
 XX
 PF 21-DEC-2001; 2001US-00023890.
 XX
 PR 21-DEC-2001; 2001US-00023890.
 XX
 PA (NOVA-) NOVAZYME PHARM INC.
 XX
 PI Canfield WM;
 XX
 DR WPI; 2003-810985/76.

DR N-PSDB; AAD62650.
 XX
 XX Producing a glycoprotein with reduced complex carbohydrates by culturing
 PT the lectin resistant mammalian cell expressing the glycoprotein for
 PT treating lysosomal storage disease.
 XX
 PS Claim 10; Page 18-19; 46pp; English.
 XX
 CC The present invention provides a method of producing a glycoprotein
 CC having reduced complex carbohydrates by culturing the lectin resistant
 CC mammalian cell expressing the glycoprotein. The method is useful for
 CC producing a glycoprotein with reduced complex carbohydrates for treating
 CC lysosomal storage disease. The present invention is also useful in gene
 CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-
 CC phosphotransferase beta subunit precursor protein
 CC
 XX
 SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 7; Length 328;
 Best Local Similarity 100.0%; Pred. No. 2,3e-160;
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHPMHMDRIVMQELQDMPPEEFDKTSFHKVRS 60
 DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAHPMHMDRIVMQELQDMPPEEFDKTSFHKVRS 60
 QY 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRLTRIHLEPLSLQDLT 120
 DB 61 EDMQAFSYFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRLTRIHLEPLSLQDLT 120
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPNLPVTKSLVYNCKRVTQKIHKAYKDK 180
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYPNLPVTKSLVYNCKRVTQKIHKAYKDK 180
 QY 181 NKYRFEIMGESEELAFKIRTNVSHVVGQLDIDKRNKRFVCINDINDHNHKAQTVKAVL 240
 DB 181 NKYRFEIMGESEELAFKIRTNVSHVVGQLDIDKRNKRFVCINDINDHNHKAQTVKAVL 240
 QY 241 RPFYESMFPPIPSQFELPREYRNFLMHLEQEWRAVRDKLKFWTHCVLATLIMFTIFSEFF 300
 DB 241 RPFYESMFPPIPSQFELPREYRNFLMHLEQEWRAVRDKLKFWTHCVLATLIMFTIFSEFF 300
 QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
 DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328

RESULT 5
 AAE25294
 ID AAE25294 standard; protein; 1196 AA.
 AC AAE25294;
 XX
 DT 30-OCT-2002 (first entry)
 XX
 DE Human nucleic acid-associated protein (NAAP-13).
 XX
 KW Human; nucleic acid-associated protein; NAAP-13; neurological disorder;
 KW arteriosclerosis; cancer; cell proliferative disorder; arteriosclerosis;
 KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
 KW autoimmune disorder; AIDS; allergy; anemia; stroke; malaria; leishmania;
 KW gene therapy; neurotropic; neuroprotective; cerebroprotective; virocidic;
 KW immunosuppressive; protozoacide; antimicrobial.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Peptide
 FT 1. 34
 FT /label= signal_peptide
 FT 1. 19
 FT /note= "Cytosolic domain"
 FT 20. 42 "Transmembrane domain"
 FT /note= "Transmembrane domain"

FT Protein 35. .1196
 FT /note="Mature human NAAP-13"
 FT Domain 43. .1152
 FT /note="Non-cytosolic domain"
 FT Domain 1153. .1175
 FT /note="Transmembrane domain"
 FT Domain 1176. .1196
 FT /note="Cytosolic domain"

XX MO200250279-A2.

XX 27-JUN-2002.

XX 19-DEC-2001; 2001WO-US050256.

XX 21-DEC-2000; 2000US-0257714P.

XX 05-JAN-2001; 2001US-0260081P.

XX 16-JAN-2001; 2001US-0262302P.

XX 23-JAN-2001; 2001US-0263823P.

XX 02-FEB-2001; 2001US-0266088P.

XX 29-OCT-2001; 2001US-0348442P.

XX (INCY-) INCYTE GENOMICS INC.

XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;
 PI Walla NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;
 PI Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;
 PI Gietzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;
 PI Lee EA, Yang J, Gorvad AE, Emerling BM, Margulis JP, Lee SY;
 PI Swarnakar A, Reddy R;

XX WPI; 2002-519887/55.
 DR N-PSDB; AAD41203.

XX Nucleic acid associated proteins and nucleic acids for diagnosing,
 PT treating and preventing cell proliferative (e.g. cancers), neurological
 PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).

XX Claim 68; Page 169-172; 193pp; English.

XX The invention relates to nucleic acid-associated proteins (NAAP) and
 CC nucleic acids. The nucleic acid and amino acid sequences are useful for
 CC diagnosing, treating and preventing cell proliferative e.g.
 CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological
 CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
 CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections
 CC (e.g. malaria, or leishmania), as well as in assessing the effects of
 CC exogenous compound on the expression of nucleic acid and amino acid
 CC sequences of nucleic acid-associated proteins. The invention is useful in
 CC gene therapy. The present sequence is human NAAP-13

XX Sequence 1196 AA;

XX Query Match 100.0%; Score 1731; DB 5; Length 1196;
 XX Best Local Similarity 100.0%; Pred. No. 1.5e-159;
 XX Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFASLSIRVKNILNSKRGFTSRKVPAMPHKIDIVNQEIQDMPPEFDTSTHKKVRS 60
 DB 869 DTFASLSIRVKNILNSKRGFTSRKVPAMPHKIDIVNQEIQDMPPEFDTSTHKKVRS 928
 QY 61 EDMQFAFSFYILMSAVOPLNISQVFDEVDVTQSGVLSDRERLTATRIHEPLSLQDLT 120
 DB 929 EDMQFAFSFYILMSAVOPLNISQVFDEVDVTQSGVLSDRERLTATRIHEPLSLQDLT 986
 QY 121 GIEHMLINCSKMLPADITQLNINPTQSGSYDNPNIPTVTKSLVTNCKEPTDKIKHAYKDX 180
 DB 989 GIEHMLINCSKMLPADITQLNINPTQSGSYDNPNIPTVTKSLVTNCKEPTDKIKHAYKDX 1048
 QY 181 NKYREINGEERIAFKMIRINSHVVGQLDIRKPKFVCLNDNIDNHKDAQTAVAYL 240
 DB 1049 NKYREINGEERIAFKMIRINSHVVGQLDIRKPKFVCLNDNIDNHKDAQTAVAYL 1108

QY 241 RDPFESMFPIPSQPELPREVRNRLFMHIEQKRAARDLCKPWTGCVLATLMTTFSPF 300
 DB 1109 RDPFESMFPIPSQPELPREVRNRLFMHIEQKRAARDLCKPWTGCVLATLMTTFSPF 1168
 QY 301 AEQILAKRKIPRRRIHKEASPNRIRV 328
 DB 1169 AEQILAKRKIPRRRIHKEASPNRIRV 1196

XX RESULT 6

XX AAE25290
 ID AAE25290 standard; protein; 1256 AA.

XX AAE25290;

XX 30-OCT-2002 (first entry)

XX Human nucleic acid-associated protein (NAAP-9).

XX Human; nucleic acid-associated protein: NAAP-9; neurological disorder;
 KW arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;
 KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
 KW autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
 KW gene therapy; neuroprotective; cerebroprotective; vitruicide;
 KW immunosuppressive; protozoacide; antimicrobial.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

CC (e.g. malaria, or leishmania), as well as in assessing the effects of
CC exogenous compound on the expression of nucleic acid and amino acid
CC sequences of nucleic acid-associated proteins. The invention is useful in
CC gene therapy. The present sequence is human NAAP-9
XX

Sequence 1256 AA;

Query Match 100.0%; Score 1731; DB 5; Length 1256;
Best Local Similarity 100.0%; Pred. No. 1.6e-159;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRHS 60
DB 929 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRHS 988
QY 61 EDMQAFASYFYILMSAVQPLNISQVFEVDVDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 989 EDMQAFASYFYILMSAVQPLNISQVFEVDVDQSGVLSDBREIRTLATRIHELPLSLQDLT 1048
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDNLPPVTKSLVTNCKEPTDKIHKAYKDK 180
DB 1049 GLEHMLINCSKMLPADITQLNINIPPTQESYDNLPPVTKSLVTNCKEPTDKIHKAYKDK 1108
QY 181 NKYRFEIMGEEELAFKMTIRTNVSHVGGQLDRIKMRKFCVCLNDINDHNDKDAQTVAVL 240
DB 1109 NKYRFEIMGEEELAFKMTIRTNVSHVGGQLDRIKMRKFCVCLNDINDHNDKDAQTVAVL 1168
QY 241 RDFYESMFPPIPSQFELPREYRNRFLLMHLEQEWRAVRDCLKFWTHCVLATLIMFTIFSFF 300
DB 1169 RDFYESMFPPIPSQFELPREYRNRFLLMHLEQEWRAVRDCLKFWTHCVLATLIMFTIFSFF 1228
QY 301 AEQILALKRKIPRRRIHKEASPNRIRV 328
DB 1229 AEQILALKRKIPRRRIHKEASPNRIRV 1256

RESULT 7

ABR61380
ID ABR61380 standard; protein; 328 AA.

ABR61380;

DT 01-AUG-2003 (first entry)

DE Murine GlcNAc-phosphotransferase beta-subunit.

KW Mouse; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;
KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;
KW lysosomal storage disease; enzyme; beta-subunit.

OS Mus musculus.

PN US6537785-B1.

PD 25-MAR-2003.

PF 10-AUG-2000; 2000US-00636077.

PR 14-SEP-1999; 99US-0153831P.

PA (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.

PI Canfield WM;

DR WPI, 2001-290356/30.

DR N-PSDB; ACC81007.

PT Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

XX
XX Disclosure; Page 35-36; 62pp; English.

CC The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has
CC nephrotropic activity, and may be useful in enzyme replacement therapy. A
CC protein of the invention (I), (II) is useful for preparing a
CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase
CC comprising a terminal mannose-6-phosphate, is useful for treating a
CC patient suffering from a lysosomal storage disease. The present sequence
CC is used in the exemplification of the invention
XX

Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.9e-157;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRHS 60
DB 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPEEFDKTSFHKVRHS 60
QY 61 EDMQAFASYFYILMSAVQPLNISQVFEVDVDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFASYFYILMSAVQPLNISQVFEVDVDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDNLPPVTKSLVTNCKEPTDKIHKAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDNLPPVTKSLVTNCKEPTDKIHKAYKDK 180
QY 181 NKYRFEIMGEEELAFKMTIRTNVSHVGGQLDRIKMRKFCVCLNDINDHNDKDAQTVAVL 240
DB 181 NKYRFEIMGEEELAFKMTIRTNVSHVGGQLDRIKMRKFCVCLNDINDHNDKDAQTVAVL 240
QY 241 RDFYESMFPPIPSQFELPREYRNRFLLMHLEQEWRAVRDCLKFWTHCVLATLIMFTIFSFF 300
DB 241 RDFYESMFPPIPSQFELPREYRNRFLLMHLEQEWRAVRDCLKFWTHCVLATLIMFTIFSFF 300
QY 301 AEQILALKRKIPRRRIHKEASPNRIRV 328
DB 301 AEQILALKRKIPRRRIHKEASPNRIRV 328

RESULT 8

ADD27818
ID ADD27818 standard; protein; 328 AA.

ADD27818;

DT 15-JAN-2004 (first entry)

DE GlcNAc-phosphotransferase associated protein #2.

KW mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase;
KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.

OS Mus musculus.

PN US2003119088-A1.

PD 26-JUN-2003.

PF 21-DEC-2001; 2001US-00023888.

PR 21-DEC-2001; 2001US-00023888.

PA (NOVA-) NOVAZYME PHARM INC.

PI Canfield W, Kudo M;

XX WPI, 2003-801323/75.

XX Phosphorylating a protein for treating a patient suffering from a
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated
PT protein.
XX
XX Disclosure; SEQ ID NO 10; 55pp; English.
XX
CC The invention relates to a method of phosphorylating a protein comprising
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-
CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of a GlcNAc-phosphotransferase associated protein.
XX
SQ Sequence 328 AA;
XX
Query Match 98.1%; Score 1698; DB 7; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.9e-157;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNTKLNKSGFTSRKVPANHPHMDIRIVMOELQDMPEEFDKTSFKVYRHS 60
DB 1 DTFADSLRYVNTKLNKSGFTSRKVPANHPHMDIRIVMOELQDMPEEFDKTSFKVYRHS 60
QY 61 EDMQFAFSFYIYMSAVOPLNISQVFEVDYDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
DB 61 EDMQFAFSFYIYMSAVOPLNISQVFEVDYDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
QY 121 GLEHMLINCSKMLPADITQNNIPTQSSYVDNLPVTKSLVTNCKEPTDKIKHAYYDK 180
DB 121 GLEHMLINCSKMLPADITQNNIPTQSSYVDNLPVTKSLVTNCKEPTDKIKHAYYDK 180
QY 181 NKYRFEIMGEELIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQIVKAVL 240
DB 181 NKYRFEIMGEELIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQIVKAVL 240
QY 241 RDFYESMPPIPSQFELPREYRNRFLMHMELQEWRAVRDCLKWTHCVLATTIIMPTIFSF 300
DB 241 RDFYESMPPIPSQFELPREYRNRFLMHMELQEWRAVRDCLKWTHCVLATTIIMPTIFSF 300
QY 301 AEQIATLKRKIPRRRIHKEASPNRIRV 328
DB 301 AEQIATLKRKIPRRRIHKEASPNRIRV 328
XX
RESULT 9
ABW01492
ID ABW01492 standard; protein; 328 AA.
XX
AC ABW01492;
XX
XX 15-JAN-2004 (first entry)
XX
DE Mouse protein #2 used to illustrate the method of the invention.
XX
XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
XX lysosomal storage disease; gastrointestinal; mouse.
XX
XX lectin resistant cell; deoxymannojirimycin; kifunensine;
XX glycosylation inhibition.
XX
XX Mus musculus.
XX
XX US2003124652-A1.
XX
XX 03-JUL-2003.
XX
XX 21-DEC-2001; 2001US-00023889.
XX
XX 21-DEC-2001; 2001US-00023889.
XX
XX (NOVA-) NOVARTIS PHARM INC.

PI Canfield WM;
XX
XX WPI; 2003-810984/76.
XX
XX Producing a high mannose glycoprotein for treating lysosomal storage
PT disease comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
XX
XX Disclosure; Page 25-26; 46pp; English.
XX
XX
CC The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is mouse protein used to
CC illustrate the method of the invention
XX
SQ Sequence 328 AA;
XX
Query Match 98.1%; Score 1698; DB 7; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.9e-157;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNTKLNKSGFTSRKVPANHPHMDIRIVMOELQDMPEEFDKTSFKVYRHS 60
DB 1 DTFADSLRYVNTKLNKSGFTSRKVPANHPHMDIRIVMOELQDMPEEFDKTSFKVYRHS 60
QY 61 EDMQFAFSFYIYMSAVOPLNISQVFEVDYDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
DB 61 EDMQFAFSFYIYMSAVOPLNISQVFEVDYDQSGVLSDBREIRTLATRIHEPLSLQDIT 120
QY 121 GLEHMLINCSKMLPADITQNNIPTQSSYVDNLPVTKSLVTNCKEPTDKIKHAYYDK 180
DB 121 GLEHMLINCSKMLPADITQNNIPTQSSYVDNLPVTKSLVTNCKEPTDKIKHAYYDK 180
QY 181 NKYRFEIMGEELIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQIVKAVL 240
DB 181 NKYRFEIMGEELIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQIVKAVL 240
QY 241 RDFYESMPPIPSQFELPREYRNRFLMHMELQEWRAVRDCLKWTHCVLATTIIMPTIFSF 300
DB 241 RDFYESMPPIPSQFELPREYRNRFLMHMELQEWRAVRDCLKWTHCVLATTIIMPTIFSF 300
QY 301 AEQIATLKRKIPRRRIHKEASPNRIRV 328
DB 301 AEQIATLKRKIPRRRIHKEASPNRIRV 328
XX
RESULT 10
ABW01541
ID ABW01541 standard; protein; 328 AA.
XX
AC ABW01541;
XX
XX 15-JAN-2004 (first entry)
XX
DE Mouse protein #2 used to illustrate the method of the invention.
XX
XX Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
XX N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse.
XX
XX Mus musculus.
XX
XX US2003124653-A1.
XX
XX 03-JUL-2003.
XX
XX 21-DEC-2001; 2001US-00023890.
XX
XX 21-DEC-2001; 2001US-00023890.

XX 21-DEC-2001; 2001US-00023890.
XX (NOVA-) NOVAZYME PHARM INC.
XX Canfield WM;
XX WPI; 2003-810985/76.
XX
XX Producing a glycoprotein with reduced complex carbohydrates by culturing
PT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.
XX
XX Disclosure; Page 25-26; 46pp; English.
XX
XX The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing the lectin resistant
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is mouse protein used to illustrate the
CC method of the invention
XX
XX Sequence 328 AA;
SQ
Query Match 98.1%; Score 1698; DB 7; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.9e-157;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKLINSKFGFTSRKVPAAHMPMHIDRIVMQELQDMFPEEFDKTSFHKVRS 60
DB 1 DTFADSLRYVNKLINSKFGFTSRKVPAAHMPMHIDRIVMQELQDMFPEEFDKTSFHKVRS 60
QY 61 EDMQAFSYYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNIPPTQESYYPNLPVTKSLVTNCKEVTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNIPPTQESYYPNLPVTKSLVTNCKEVTDKIKHAYKDK 180
QY 181 NKXRFELMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNNKDAQTVAVL 240
DB 181 NKXRFELMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNNKDAQTVAVL 240
QY 241 RDEFSWMPFIPSOQFELPREYRNRFILMHLEQEMRAVRDKLKFMTGCVLDTLITFISFF 300
DB 241 RDEFSWMPFIPSOQFELPREYRNRFILMHLEQEMRAVRDKLKFMTGCVLDTLITFISFF 300
QY 301 AEQIALLKKRIKIPRRRIHKEASPNRIKRV 328
DB 301 AEQIALLKKRIKIPRRRIHKEASPNRIKRV 328
RESULT 11
ADD27810
ID ADD27810 standard; protein; 1199 AA.
XX
XX ADD27810;
AC
XX
XX 15-JAN-2004 (first entry)
DT
XX
XX Soluble human GlcNAc-6-phosphotransferase.
DE
XX
XX human; protein phosphorylation; soluble GlcNAc-6-phosphotransferase;
KM UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease;
KM enzyme.
XX
XX Synthetic.
OS Homo sapiens.
OS
XX
XX US2003119088-A1.
XX

PD 26-JUN-2003.
XX
XX 21-DEC-2001; 2001US-00023888.
PF
XX
XX 21-DEC-2001; 2001US-00023888.
PR
XX
XX (NOVA-) NOVAZYME PHARM INC.
XX
XX Canfield W, Kudo M;
PI
XX
XX WPI; 2003-801323/75.
DR
XX
XX N-PSDB; ADD27809.
DR
XX
XX Phosphorylating a protein for treating a patient suffering from a
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein
PT with a soluble GlcNAc-6-phosphotransferase and producing a phosphorylated
PT protein.
XX
XX Claim 3; SEQ ID NO 2; 55pp; English.
XX
XX The invention relates to a method of phosphorylating a protein comprising
CC contacting the protein with a soluble GlcNAc-6-phosphotransferase (UDP-N-
CC acetylglucosamine) and producing a phosphorylated protein. The method is
CC useful for treating a patient suffering from a lysosomal storage disease
CC e.g. Fabry's disease. The present sequence represents the amino acid
CC sequence of soluble human GlcNAc-6-phosphotransferase.
XX
XX Sequence 1199 AA;
SQ
Query Match 85.9%; Score 1487; DB 7; Length 1199;
Best Local Similarity 100.0%; Pred. No. 1.1e-135;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKLINSKFGFTSRKVPAAHMPMHIDRIVMQELQDMFPEEFDKTSFHKVRS 60
DB 919 DTFADSLRYVNKLINSKFGFTSRKVPAAHMPMHIDRIVMQELQDMFPEEFDKTSFHKVRS 978
QY 61 EDMQAFSYYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 979 EDMQAFSYYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHELPLSLQDLT 1038
QY 121 GLEHMLINCSKMLPADITQLNIPPTQESYYPNLPVTKSLVTNCKEVTDKIKHAYKDK 180
DB 1039 GLEHMLINCSKMLPADITQLNIPPTQESYYPNLPVTKSLVTNCKEVTDKIKHAYKDK 1098
QY 181 NKXRFELMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNNKDAQTVAVL 240
DB 1099 NKXRFELMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNNKDAQTVAVL 1158
QY 241 RDEFSWMPFIPSOQFELPREYRNRFILMHLEQEMRAVRDKL 281
DB 1159 RDEFSWMPFIPSOQFELPREYRNRFILMHLEQEMRAVRDKL 1199
RESULT 12
ABW01487
ID ABW01487 standard; protein; 1199 AA.
XX
XX ABW01487;
AC
XX
XX 15-JAN-2004 (first entry)
DT
XX
XX N-acetylglucosamine-1(GlcNAc)-6-phosphotransferase.
DE
XX
XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell;
KM lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;
KM gastrointestinal; enzyme; lectin resistant cell; deoxymannojirimycin;
KM kifunensine; glycosylation inhibition.
XX
XX Unidentified.
OS
XX
XX US2003124652-A1.
XX

PD 03-JUL-2003.
XX
XX 21-DEC-2001; 2001US-00023899.
XX
XX 21-DEC-2001; 2001US-00023899.
XX
XX (NOVA-) NOVAZYME PHARM INC.
XX
XX Canfield WM;
XX PI
XX DR WPI; 2003-810984/76.
XX DR N-PSDB; AAD62490.
XX
PT Producing a high mannose glycoprotein for treating lysosomal storage
PT disease, comprises culturing the lectin resistant mammalian cell in the
PT presence of deoxymannojirimycin and kifunensine.
XX
XX Claim 8; Page 10-13; 46pp; English.
XX
CC The invention relates to a method for producing a high mannose
CC glycoprotein. The method comprises: introducing and expressing a
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing
CC the cell in the presence of a lectin to obtain a lectin resistant cell;
CC isolating the cell; culturing the cell in the presence of
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the
CC glycoprotein; and collecting the glycoprotein. The invention is useful in
CC gene therapy. The method is useful for producing a high mannose
CC glycoprotein in a complex carbohydrate deficient cell for treating
CC lysosomal storage disease. The present sequence is N-acetylglucosamine-1
CC (GlcNAc)-phosphotransferase
XX
SQ Sequence 1199 AA;
Query Match 85.9%; Score 1487; DB 7; Length 1199;
Best Local Similarity 100.0%; Pred. No. 1.1e-135;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAAHFMHIDRIWQELQDMFPEEFDKTSFHKVRS 60
DB 919 DTFADSLRYVKNILNSKFGFTSRKVPAAHFMHIDRIWQELQDMFPEEFDKTSFHKVRS 978
QY 61 EDMQPAFSFYIYMSAVOPLNISQVFDEVDTQSGVLSDBREIRTLATRIHELPLSLQDIT 120
DB 979 EDMQPAFSFYIYMSAVOPLNISQVFDEVDTQSGVLSDBREIRTLATRIHELPLSLQDIT 1038
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTDKIKHAYXDK 180
DB 1039 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTDKIKHAYXDK 1098
QY 181 NKYREIMGEEBEIAFMKIRTNVSHVGGQLDIRKPKRFVCCLNDINDHMKDAQTVAVL 240
DB 1099 NKYREIMGEEBEIAFMKIRTNVSHVGGQLDIRKPKRFVCCLNDINDHMKDAQTVAVL 1158
QY 241 RDFYESMFPPIPSQFELPREYRNRFLHMHLOQEMRAYRDXLK 281
DB 1159 RDFYESMFPPIPSQFELPREYRNRFLHMHLOQEMRAYRDXLK 1199
RESULT 13
ABW01536
ID ABW01536 standard; protein; 1199 AA.
XX
XX ABW01536;
XX
XX 15-JAN-2004 (first entry)
XX
XX N-acetylglucosamine-1 (GlcNAc)-phosphotransferase protein.
XX
XX Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
XX N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme.
XX
XX Unidentified.
XX

PN US2003124653-A1.
XX
XX 03-JUL-2003.
XX
XX 21-DEC-2001; 2001US-00023890.
XX
XX 21-DEC-2001; 2001US-00023890.
XX
XX (NOVA-) NOVAZYME PHARM INC.
XX
XX Canfield WM;
XX PI
XX DR WPI; 2003-810984/76.
XX DR N-PSDB; AAD62490.
XX
PT Producing a glycoprotein with reduced complex carbohydrates by culturing
PT the lectin resistant mammalian cell expressing the glycoprotein for
PT treating lysosomal storage disease.
XX
XX Claim 8; Page 10-13; 46pp; English.
XX
CC The present invention provides a method of producing a glycoprotein
CC having reduced complex carbohydrates by culturing the lectin resistant
CC mammalian cell expressing the glycoprotein. The method is useful for
CC producing a glycoprotein with reduced complex carbohydrates for treating
CC lysosomal storage disease. The present invention is also useful in gene
CC therapy. The present sequence is N-acetylglucosamine (GlcNAc)-
CC phosphotransferase protein
XX
SQ Sequence 1199 AA;
Query Match 85.9%; Score 1487; DB 7; Length 1199;
Best Local Similarity 100.0%; Pred. No. 1.1e-135;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAAHFMHIDRIWQELQDMFPEEFDKTSFHKVRS 60
DB 919 DTFADSLRYVKNILNSKFGFTSRKVPAAHFMHIDRIWQELQDMFPEEFDKTSFHKVRS 978
QY 61 EDMQPAFSFYIYMSAVOPLNISQVFDEVDTQSGVLSDBREIRTLATRIHELPLSLQDIT 120
DB 979 EDMQPAFSFYIYMSAVOPLNISQVFDEVDTQSGVLSDBREIRTLATRIHELPLSLQDIT 1038
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTDKIKHAYXDK 180
DB 1039 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTDKIKHAYXDK 1098
QY 181 NKYREIMGEEBEIAFMKIRTNVSHVGGQLDIRKPKRFVCCLNDINDHMKDAQTVAVL 240
DB 1099 NKYREIMGEEBEIAFMKIRTNVSHVGGQLDIRKPKRFVCCLNDINDHMKDAQTVAVL 1158
QY 241 RDFYESMFPPIPSQFELPREYRNRFLHMHLOQEMRAYRDXLK 281
DB 1159 RDFYESMFPPIPSQFELPREYRNRFLHMHLOQEMRAYRDXLK 1199
RESULT 14
ABU07381
ID ABU07381 standard; protein; 1459 AA.
XX
XX ABU07381;
XX
XX 28-JAN-2003 (first entry)
XX
XX Human protein NOV12.
XX
XX Human; NOVX; cardiomyopathy; atherosclerosis; cancer; hypertension;
XX diabetes; inflammation; autoimmune disorder; allergy; blood disorder;
XX acquired immunodeficiency syndrome; AIDS; obesity; asthma;
XX immunoglobulin A nephropathy; cirrhosis; arthritis; Alzheimer's disease;
XX Parkinson's disease; gout; infection; stroke; muscular dystrophy;
XX epilepsy; wasting disorder; neurogenesis; cell differentiation;
XX cell proliferation; haematopoiesis; wound healing; angiogenesis;
XX

XX gene therapy; single nucleotide polymorphism; SNP.
XX Homo sapiens.
XX WO200265922-A2.
XX 31-OCT-2002.
XX 11-APR-2002; 2002WO-US011634.
XX 23-APR-2001; 2001US-0285748P.
XX 24-APR-2001; 2001US-0286068P.
XX 25-APR-2001; 2001US-0286292P.
XX 03-MAY-2001; 2001US-0288334P.
XX 16-MAY-2001; 2001US-0291241P.
XX 14-SEP-2001; 2001US-0322284P.
XX (CURA-) CUREGEN CORP.
XX (MILL-) MILLENNIUM PHARM INC.
XX Pena CE, Guo X, Shinkets RA, Padigaru M, Kekuda R, Spytek KA,
XX Mehrahan F, Topper UN, Malyankar UM, Wasserman S, Edinger S,
XX Smithson G, Gunther E, Komuves L,
XX WPI; 2003-058712/05.
XX N-PSDB; ABX10235.
XX New NOXV polypeptides and nucleic acids, useful for preventing or
XX treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,
XX atherosclerosis or diabetes, and in chromosome mapping, tissue typing or
XX pharmacogenomics.
XX Claim 1, Page 94-95; 301pp; English.
XX The invention relates to an isolated polypeptide comprising any one of 17
XX human NOXV (1-9, 10a, 10b, 11-16) appearing as AB007369-AB007385, a
XX mature form of it, or a variant of them, where one or more residues of
XX the variant differs in not more than 15 % from the residues of the
XX sequence of them and their encoding polynucleotides appearing as ABX10223
XX -ABX10239. Also included are NOXV expression vectors, transformed cells,
XX antibodies, identifying an agent that binds to or modulates the
XX expression or activity of NOXV and screening for a modulator of activity
XX or of latency or predisposition to a NOXV-associated disorder. The NOXV
XX polypeptides, polynucleotides and antibodies are useful in manufacturing
XX a medicament for treating or preventing a syndrome associated with NOXV-
XX associated disorder, such as cardiomyopathy, atherosclerosis, cancer,
XX hypertension, diabetes, inflammation, autoimmune disorders, allergies,
XX blood disorders, acquired immunodeficiency syndrome (AIDS), obesity,
XX asthma, immunoglobulin (Ig)A nephropathy, cirrhosis, arthritis,
XX Alzheimer's disease, Parkinson's disease, goitre, infections (e.g.
XX bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy, and
XX other wasting disorders associated with chronic diseases. The nucleic
XX acids and polypeptides may also be used as targets for the identification
XX of small molecules that modulate or inhibit e.g. neurogenesis, cell
XX differentiation, cell proliferation, haematopoiesis, wound healing and
XX angiogenesis, in gene therapy, in generation of antibodies that bind
XX immunospecifically to NOXV substances for use in therapeutic or
XX diagnostic methods. The nucleic acids are further used as hybridisation
XX probes, in chromosome mapping, tissue typing, preventive medicine, and
XX pharmacogenomics. The polypeptides are also useful as vaccines. The
XX present sequence represents a NOXV protein
XX
XX Sequence 1459 AA;
XX
XX Query Match 65.9%; Score 1140; DB 6; Length 1459;
XX Best Local Similarity 98.6%; Pred. No. 1.2e-101;
XX Matches 218; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 DFFPASLRVYNKILNSKGFTRSRKVPANMHPMIDINVOCELODMPEEFDKTSFPRKXHS 60
XX |||
XX 1209 DFFPASLRVYNKILNSKGFTRSRKVPANMHPMIDINVOCELODMPEEFDKTSFPRKXHS 1268
XX
XX 61 EDMQFAFSFYIYMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDILT 120

DB 1269 EDMQFAFSFYIYMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDILT 1328
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVTKSLVTWNCSPYTDKIKKAYKDK 180
DB 1329 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVTKSLVTWNCSPYTDKIKKAYKDK 1386
QY 181 NKYRFEIMGEEIEAFPMIRTNVSHVVGQDDIRKPNRKEVC 221
DB 1389 NKYRFEIMGEEIEAFPMIRTNVSHVVGQDDIRKPNRISLC 1429
RESULT 15
AAB93369 standard; protein, 132 AA.
ID AAB93369
AC AAB93369;
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:12517.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EPI074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX
XX Claim 8; SEQ ID NO 12517; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
XX represent human amino acid sequences; and AAH13629 to AAH13633 represent
XX oligonucleotides, all of which are used in the exemplification of the

CC Present invention
XX
SQ Sequence 132 AA;

Query Match 40.7%; Score 704; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.4e-60;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 MIRTNVSHVVGQLDDIRKNPRKFCVINDNIDHNHKAQTVKAVLRDPFESMFPISQPEL 256
DB 1 MIRTNVSHVVGQLDDIRKNPRKFCVINDNIDHNHKAQTVKAVLRDPFESMFPISQPEL 60

QY 257 PREYRNRFLMHHELOEMRAYRDCLKFTWTCVLTATIMFTTISFFFAEQLIALKRKIFPRRR 316
DB 61 PREYRNRFLMHHELOEMRAYRDCLKFTWTCVLTATIMFTTISFFFAEQLIALKRKIFPRRR 120

QY 317 IHKEASPNRIRV 328
DB 121 IHKEASPNRIRV 132

Search completed: July 26, 2004, 11:09:10
Job time : 32.2575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:06:34 ; Search time 9.24536 Seconds

(without alignments)
1831.549 Million cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731

Sequence: 1 DTFADSLRYVKNKILNSKRGF.....RKIPRRRIHKEASPNRIRV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 segs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*

2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*

3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*

4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*

5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*

6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1731	100.0	328	4	US-09-635-872A-2
2	1731	100.0	328	4	US-09-636-077A-2
3	1731	100.0	328	4	US-09-636-060C-2
4	1731	100.0	328	4	US-09-986-552-2
5	1698	98.1	328	4	US-09-635-872A-8
6	1698	98.1	328	4	US-09-636-077A-8
7	1698	98.1	328	4	US-09-636-060C-8
8	1698	98.1	328	4	US-09-986-552-8
9	402	23.2	502	4	US-09-635-872A-13
10	402	23.2	502	4	US-09-636-077A-13
11	402	23.2	502	4	US-09-636-060C-13
12	402	23.2	502	4	US-09-986-552-13
13	145	8.4	28	4	US-09-635-872A-27
14	145	8.4	28	4	US-09-636-077A-27
15	145	8.4	28	4	US-09-636-060C-27
16	145	8.4	28	4	US-09-986-552-27
17	113.5	6.6	545	4	US-08-936-107A-10
18	101	5.8	2662	4	US-09-595-684B-31
19	100.5	5.8	259	3	US-08-469-318-137
20	100.5	5.8	259	3	US-08-469-318-155
21	100.5	5.8	259	3	US-08-468-609A-137
22	100.5	5.8	259	3	US-08-468-609A-155
23	100.5	5.8	259	4	US-08-446-872A-137
24	100.5	5.8	259	4	US-08-446-872A-155
25	100.5	5.8	259	4	US-08-762-227A-137
26	100.5	5.8	259	4	US-08-762-227A-155
27	100.5	5.8	259	5	PCT-US95-01185-137

28	100.5	5.8	259	5	PCT-US95-01185-155	Sequence 155, App
29	96.5	5.6	588	4	US-09-519-232-2	Sequence 2, Appli
30	93.5	5.4	652	4	US-08-559-896B-2	Sequence 2, Appli
31	93.5	5.4	652	4	US-09-351-794A-2	Sequence 2, Appli
32	93.5	5.4	1199	3	US-09-308-742-2	Sequence 2, Appli
33	93.5	5.4	1199	4	US-09-332-295-4	Sequence 4, Appli
34	93.5	5.4	1199	4	US-09-709-979-4	Sequence 4, Appli
35	93.5	5.4	1199	4	US-10-147-268-4	Sequence 4, Appli
36	93.5	5.4	2184	4	US-09-417-485D-6	Sequence 6, Appli
37	92.5	5.3	696	4	US-09-107-532A-5205	Sequence 5205, Ap
38	92	5.3	315	4	US-09-584-568C-2	Sequence 2, Appli
39	91.5	5.3	322	3	US-08-469-318-149	Sequence 149, App
40	91.5	5.3	322	3	US-08-469-318-156	Sequence 156, App
41	91.5	5.3	322	3	US-08-469-318-157	Sequence 157, App
42	91.5	5.3	322	3	US-08-468-609A-149	Sequence 149, App
43	91.5	5.3	322	3	US-08-468-609A-156	Sequence 156, App
44	91.5	5.3	322	3	US-08-468-609A-157	Sequence 157, App
45	91.5	5.3	322	4	US-08-446-872A-149	Sequence 149, App

ALIGNMENTS

```
RESULT 1
US-09-635-872A-2
; Sequence 2, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613050
; CURRENT APPLICATION NUMBER: US/09/635,872A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-635-872A-2
```

Query Match 100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.7e-169;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DTFADSLRYVKNKILNSKRGFTSRKVPANHPHMDIRIVQELDMPPEEFDTKSFHKVRS	60
DB	1	DTFADSLRYVKNKILNSKRGFTSRKVPANHPHMDIRIVQELDMPPEEFDTKSFHKVRS	60
QY	61	EDMQAFSYFYVYMSAVQPLNISQVFEVDVTDQSGVLSREIRRTATRIHEPLSLQDIT	120
DB	61	EDMQAFSYFYVYMSAVQPLNISQVFEVDVTDQSGVLSREIRRTATRIHEPLSLQDIT	120
QY	121	GLEHMTLINSKMLPATITQLNIPPTQESYVQNPVPPVKSIVTNCKPVTDXHKYKXK	180
DB	121	GLEHMTLINSKMLPATITQLNIPPTQESYVQNPVPPVKSIVTNCKPVTDXHKYKXK	180
QY	181	NKTRFELMGEELAFQKIRITNVSHVVGQDDIRKKNRKEVCINDINDHKNKDAQVKA	240
DB	181	NKTRFELMGEELAFQKIRITNVSHVVGQDDIRKKNRKEVCINDINDHKNKDAQVKA	240
QY	241	RDTYESMFPDPSQFELPREYRNFFLMHQLQEMRAYRDCKKPTWTCVATLIMFTTFSFP	300
DB	241	RDTYESMFPDPSQFELPREYRNFFLMHQLQEMRAYRDCKKPTWTCVATLIMFTTFSFP	300
QY	301	AEQILAKRKIFPRRIHKEASPNRIRV 328	
DB	301	AEQILAKRKIFPRRIHKEASPNRIRV 328	

```
US-09-636-077A-2
; Sequence 2, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-636-077A-2

Query Match      100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.7e-169;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKILNSKGFSTRKVPAAHMHMIDRIYVQELQDMFPEEDKTSFHKVRS 60
DB 1 DTFADSLRYVNKILNSKGFSTRKVPAAHMHMIDRIYVQELQDMFPEEDKTSFHKVRS 60
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
QY 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
DB 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
QY 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
DB 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
QY 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
QY 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328

RESULT 3
US-09-636-060C-2
; Sequence 2, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US00CNT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-636-060C-2

Query Match      100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.7e-169;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKILNSKGFSTRKVPAAHMHMIDRIYVQELQDMFPEEDKTSFHKVRS 60
DB 1 DTFADSLRYVNKILNSKGFSTRKVPAAHMHMIDRIYVQELQDMFPEEDKTSFHKVRS 60
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
QY 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
DB 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
QY 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
DB 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
QY 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
QY 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
```

```
DB 1 DTFADSLRYVNKILNSKGFSTRKVPAAHMHMIDRIYVQELQDMFPEEDKTSFHKVRS 60
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
QY 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
DB 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
QY 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
DB 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
QY 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
QY 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328

RESULT 4
US-09-986-552-2
; Sequence 2, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77D1V
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-552-2

Query Match      100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.7e-169;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKILNSKGFSTRKVPAAHMHMIDRIYVQELQDMFPEEDKTSFHKVRS 60
DB 1 DTFADSLRYVNKILNSKGFSTRKVPAAHMHMIDRIYVQELQDMFPEEDKTSFHKVRS 60
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDITDQSGVLSDBREIRTLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHAYKDX 180
QY 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFMIRTNVSHVVGQLDIRKPKRFVCINDNIDHNHDAQTVKAVL 240
QY 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
DB 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
QY 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
DB 241 RDFYESMFPISQFELPREYRNRFLMHMLQEMRAYRDKLKFWTHCVLATTLMFTIFSFF 300
QY 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
QY 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
DB 301 AEQLIALKRIKIFPRRRIHKEASPNRIRV 328
```


RESULT 5
US-09-635-872A-8
; Sequence 8, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613050
; CURRENT APPLICATION NUMBER: US/09/635, 872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153, 831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-635-872A-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 2.3e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAMPHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
DB 1 DTFADSLRYVNKILNSKFGFTSRKVPAMPHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
QY 61 EDMQAFASYFYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
DB 61 EDMQAFASYFYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
QY 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKDK 180
QY 181 NKYREIMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTVKAYVL 240
DB 181 NKYREIMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTVKAYVL 240
QY 241 RDPYESMFPISQFELPREYRNRFILMHLEQEMRAYRDKLKFWTHCVLATLIIFTIFSFF 300
DB 241 RDPYESMFPISQFELPREYRNRFILMHLEQEMRAYRDKLKFWTHCVLATLIIFTIFSFF 300
QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328

RESULT 6
US-09-636-077A-8
; Sequence 8, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636, 077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153, 831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-077A-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 2.3e-165;

Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAMPHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
DB 1 DTFADSLRYVNKILNSKFGFTSRKVPAMPHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
QY 61 EDMQAFASYFYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
DB 61 EDMQAFASYFYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
QY 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKDK 180
QY 181 NKYREIMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTVKAYVL 240
DB 181 NKYREIMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTVKAYVL 240
QY 241 RDPYESMFPISQFELPREYRNRFILMHLEQEMRAYRDKLKFWTHCVLATLIIFTIFSFF 300
DB 241 RDPYESMFPISQFELPREYRNRFILMHLEQEMRAYRDKLKFWTHCVLATLIIFTIFSFF 300
QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328

RESULT 7
US-09-636-060C-8
; Sequence 8, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119USOCONT
; CURRENT APPLICATION NUMBER: US/09/636, 060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153, 831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-8

Query Match 98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 2.3e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAMPHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
DB 1 DTFADSLRYVNKILNSKFGFTSRKVPAMPHMIDRIWMQELQDMPEEFDKTSFHKVRHS 60
QY 61 EDMQAFASYFYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
DB 61 EDMQAFASYFYIYMSAVQPLNISQVPEVDVTDQSGVLSDBREIRTLATRIHEPLSLQDPLT 120
QY 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKDK 180
DB 121 GLEHMLINCSKMLPADITQANNIPPTQESYDPLNPVTYKSLVTNCKPVTDKIKHAYKDK 180
QY 181 NKYREIMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTVKAYVL 240
DB 181 NKYREIMGEEBEIAFKMIRTNVSHVVGQLDIRKPKRFVCLNDNDHNHKAQTVKAYVL 240
QY 241 RDPYESMFPISQFELPREYRNRFILMHLEQEMRAYRDKLKFWTHCVLATLIIFTIFSFF 300
DB 241 RDPYESMFPISQFELPREYRNRFILMHLEQEMRAYRDKLKFWTHCVLATLIIFTIFSFF 300
QY 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328
DB 301 AEQILAKRKIFPRRRIHKEASPNRIRV 328

```
Db      301 AEQIILAKRKIFPPRRIRHKASPDRIIV 328

RESULT 8
US-09-986-552-8
; Sequence 8, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US72DIY
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-552-8

Query Match      98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 2.3e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY      1 DTFADSLRYVKNILNSKFGFTSRKYPAPHPMHIDRIWQELQDMPEEFEDKTSFHKVHS 60
Db      1 DTFADSLRYVKNILNSKFGFTSRKYPAPHPMHIDRIWQELQDMPEEFEDKTSFHKVHS 60
QY      61 EDMQAFSYFYIYMSAVOPLNISQVPEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
Db      61 EDMQAFSYFYIYMSAVOPLNISQVPEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
QY      121 GLEHMLINCSKMLPADITQLNIPPTQESYDDPNLPVTKSLVTVCKEVTDKIKHAYDK 180
Db      121 GLEHMLINCSKMLPADITQLNIPPTQESYDDPNLPVTKSLVTVCKEVTDKIKHAYDK 180
QY      181 NKRYPEIMGEELIAFKMIRTNVSHVVGQLDIRKNPKKFCVCLNDINDHNDKDAQTAVL 240
Db      181 NKRYPEIMGEELIAFKMIRTNVSHVVGQLDIRKNPKKFCVCLNDINDHNDKDAQTAVL 240
QY      241 RPYESMPPIPSQFELPRYRRLAHMHELOEMRAYRDKLKEWTHCTLATLIMFTISFF 300
Db      241 RPYESMPPIPSQFELPRYRRLAHMHELOEMRAYRDKLKEWTHCTLATLIMFTISFF 300
QY      301 AEQIILAKRKIFPPRRIRHKASPDRIIV 328
Db      301 AEQIILAKRKIFPPRRIRHKASPDRIIV 328

RESULT 9
US-09-635-872A-13
; Sequence 13, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 19613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-635-872A-13

Query Match      23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.3e-32;
Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY      1 DTFADSLRYVKNILNSKFGFTSRKYPAPHPMHIDRIWQELQDMPEEFEDKTSFHKVHS 60
Db      259 DIYSHSLIATNMILNAYGFKARHVLAVHGFLLIDDIYAMORRHHQIILDAHORFRAP 318
QY      61 EDMQAFSYFYIYMSAVOPLNISQVPEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
Db      319 TDQVAFAYYSPLMSETKWMSVEIFDEFDTGSAVSRVETFLTRIYQEPDLSAMR 378
QY      121 GLEHMLINCSK---MLPADITQLNIPPTQESYDDPNLPVTKSLVTVCKEVTDKIKHA 176
Db      379 YFEENVQNCTRNIGMLKVDIVESHSL--VYERYEDSNLPITTRDLVVCPLLAELAAAN 436
QY      177 YKDNKRPFIOME--EELAFKMIKIRTNVSHVVGQLDIRKNPKKFCVCLNDINDHN 229
Db      437 FAVRPKTNHVSFKRISHSNFMILSNLTVEVSLDRLRNPKKFCVCLNDINDAN 491

RESULT 10
US-09-636-077A-13
; Sequence 13, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 19612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-636-077A-13

Query Match      23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.3e-32;
Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY      1 DTFADSLRYVKNILNSKFGFTSRKYPAPHPMHIDRIWQELQDMPEEFEDKTSFHKVHS 60
Db      259 DIYSHSLIATNMILNAYGFKARHVLAVHGFLLIDDIYAMORRHHQIILDAHORFRAP 318
QY      61 EDMQAFSYFYIYMSAVOPLNISQVPEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120
Db      319 TDQVAFAYYSPLMSETKWMSVEIFDEFDTGSAVSRVETFLTRIYQEPDLSAMR 378
QY      121 GLEHMLINCSK---MLPADITQLNIPPTQESYDDPNLPVTKSLVTVCKEVTDKIKHA 176
Db      379 YFEENVQNCTRNIGMLKVDIVESHSL--VYERYEDSNLPITTRDLVVCPLLAELAAAN 436
QY      177 YKDNKRPFIOME--EELAFKMIKIRTNVSHVVGQLDIRKNPKKFCVCLNDINDHN 229
Db      437 FAVRPKTNHVSFKRISHSNFMILSNLTVEVSLDRLRNPKKFCVCLNDINDAN 491

RESULT 11
US-09-636-060C-13
; Sequence 13, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US0CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
```

PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.1
SEQ ID NO 13
LENGTH: 502
TYPE: PRF
ORGANISM: Drosophila melanogaster
US-09-636-060C-13

Query Match 23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.3e-32;
Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAMPHMIDRIWQELQDMPEEFDKTSFHKVHS 60
DB 259 DYSLSLATNMILNRAVGFARHVAHVGFLIDQIVBAQRRHQIILDTAQRFRAP 318
QY 61 EDMQAFSYFYIYLSAVOPLNISQVDEVDTDQSGVLSDRERITLRIHEPLSLQDLT 120
DB 319 TDLQYAFAYYSFLMSEETKMSVEEIEFDEPDIGSATWSDREVRPFLTRIYQPLDWSMR 378
QY 121 GLEHMLINCSK---MLPADITQJLNNIPTQESYVDPLPVYTSKLVNCKSPYDKIHA 176
DB 379 YFEENVQCTNRLGMLKVDYVESHSTL--VYERYEDSNLPITTRDLVVRCPDLAEALAN 436
QY 177 YDKKKYREIIMG--EEIAFQIRTNVSHVVGQLDIRKNPKFVCINDNIDHN 229
DB 437 FAVRPKNVHVPKRTSHSNFMULTSNLVEVESLDRLRNPKKNCINDNIDAN 491

RESULT 12
US-09-986-552-13
Sequence 13, Application US/09986552
Patent No. 6670165
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 215089US77DIV
CURRENT APPLICATION NUMBER: US/09/986,552
CURRENT FILING DATE: 2001-11-09
PRIOR APPLICATION NUMBER: 09/635,872
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.1
SEQ ID NO 13
LENGTH: 502
TYPE: PRF
ORGANISM: Drosophila melanogaster
US-09-986-552-13

Query Match 23.2%; Score 402; DB 4; Length 502;
Best Local Similarity 34.5%; Pred. No. 1.3e-32;
Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAMPHMIDRIWQELQDMPEEFDKTSFHKVHS 60
DB 259 DYSLSLATNMILNRAVGFARHVAHVGFLIDQIVBAQRRHQIILDTAQRFRAP 318
QY 61 EDMQAFSYFYIYLSAVOPLNISQVDEVDTDQSGVLSDRERITLRIHEPLSLQDLT 120
DB 319 TDLQYAFAYYSFLMSEETKMSVEEIEFDEPDIGSATWSDREVRPFLTRIYQPLDWSMR 378
QY 121 GLEHMLINCSK---MLPADITQJLNNIPTQESYVDPLPVYTSKLVNCKSPYDKIHA 176
DB 379 YFEENVQCTNRLGMLKVDYVESHSTL--VYERYEDSNLPITTRDLVVRCPDLAEALAN 436
QY 177 YDKKKYREIIMG--EEIAFQIRTNVSHVVGQLDIRKNPKFVCINDNIDHN 229
DB 437 FAVRPKNVHVPKRTSHSNFMULTSNLVEVESLDRLRNPKKNCINDNIDAN 491

RESULT 13
US-09-635-872A-27
Sequence 27, Application US/09635872A
Patent No. 6534300
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 195612US0
CURRENT APPLICATION NUMBER: US/09/635,872A
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.0
SEQ ID NO 27
LENGTH: 28
TYPE: PRF
ORGANISM: Bos taurus
US-09-635-872A-27

Query Match 8.4%; Score 145; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAM 28
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAM 28
RESULT 14
US-09-636-077A-27
Sequence 27, Application US/09636077A
Patent No. 6537785
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM
TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
FILE REFERENCE: 195612US0
CURRENT APPLICATION NUMBER: US/09/636,077A
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.0
SEQ ID NO 27
LENGTH: 28
TYPE: PRF
ORGANISM: Bos taurus
US-09-636-077A-27

Query Match 8.4%; Score 145; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAM 28
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAM 28
RESULT 15
US-09-636-060C-27
Sequence 27, Application US/09636060C
Patent No. 6642038
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM M
TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
FILE REFERENCE: 210119US0CONT
CURRENT APPLICATION NUMBER: US/09/636,060C
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.1

; SEQ ID NO 27
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-636-060C-27

Query Match 8.4%; Score 145; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAH 28
|||
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAAH 28

Search completed: July 26, 2004, 11:13:19
Job time : 10.2454 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:11:44 ; Search time 25.2146 Seconds

(without alignments)
4073.614 Million cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731
Sequence: 1 DTFADSLRYNKKINSKFGF.....RKIFPRRIHKEASPNRIRV 328

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US50_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1731	100.0	328	9	US-09-895-072-2
2	1731	100.0	328	9	US-09-895-072-2
3	1731	100.0	328	14	US-10-023-888-5
4	1731	100.0	328	14	US-10-023-888-5
5	1731	100.0	328	14	US-10-023-889-5
6	1731	100.0	328	14	US-10-023-890-5
7	1731	100.0	328	14	US-10-024-197-5
8	1731	100.0	328	14	US-10-023-894-5
9	1731	100.0	328	14	US-10-306-686-2
10	1698	98.1	328	9	US-09-895-072-8
11	1698	98.1	328	9	US-09-895-072-8
12	1698	98.1	328	14	US-10-023-888-10
13	1698	98.1	328	14	US-10-023-889-10
14	1698	98.1	328	14	US-10-023-890-10
15	1698	98.1	328	14	US-10-024-197-10
15	1698	98.1	328	14	US-10-023-894-10

16	1698	98.1	328	14	US-10-306-686-8	Sequence 8, Appli
17	1487	85.9	1199	14	US-10-023-888-2	Sequence 2, Appli
18	1487	85.9	1199	14	US-10-023-889-2	Sequence 2, Appli
19	1487	85.9	1199	14	US-10-023-890-2	Sequence 2, Appli
20	1487	85.9	1199	14	US-10-024-197-2	Sequence 2, Appli
21	1487	85.9	1199	14	US-10-023-894-2	Sequence 2, Appli
22	1290	74.5	248	15	US-10-120-801-90	Sequence 90, Appli
23	1140	65.9	663	15	US-10-120-801-87	Sequence 87, Appli
24	1140	65.9	1459	15	US-10-120-801-26	Sequence 26, Appli
25	581	33.6	384	15	US-10-120-801-89	Sequence 89, Appli
26	505	29.2	652	15	US-10-120-801-91	Sequence 91, Appli
27	402	23.2	502	9	US-09-895-072-13	Sequence 13, Appli
28	402	23.2	502	9	US-09-895-072-13	Sequence 13, Appli
29	402	23.2	502	14	US-10-023-888-16	Sequence 16, Appli
30	402	23.2	502	14	US-10-023-889-16	Sequence 16, Appli
31	402	23.2	502	14	US-10-023-890-16	Sequence 16, Appli
32	402	23.2	502	14	US-10-024-197-16	Sequence 16, Appli
33	402	23.2	502	14	US-10-023-894-16	Sequence 16, Appli
34	402	23.2	502	14	US-10-306-686-13	Sequence 30775, A
35	300	17.3	55	14	US-10-023-386-30775	Sequence 349, App
36	228	13.2	112	13	US-10-079-623-349	Sequence 27, Appli
37	145	8.4	28	9	US-09-895-072-27	Sequence 27, Appli
38	145	8.4	28	9	US-09-895-072-27	Sequence 27, Appli
39	145	8.4	28	14	US-10-306-686-27	Sequence 27, Appli
40	124	7.2	1312	15	US-10-369-493-1950	Sequence 1950, Ap
41	112	6.5	46	14	US-10-437-963-160369	Sequence 160369,
42	110.5	6.4	1015	16	US-10-437-963-160369	Sequence 160369,
43	104.5	6.0	937	16	US-10-437-963-160445	Sequence 160445,
44	104.5	6.0	1999	15	US-10-028-248A-107	Sequence 107, App
45	104.5	6.0	1999	15	US-10-107-782-107	Sequence 107, App

ALIGNMENTS

RESULT 1
US-09-895-072-2
Sequence 2, Application US/09895072
Patent No. US20020025550A1
GENERAL INFORMATION:
APPLICANT: CANFIELD, WILLIAM M
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
FILE REFERENCE: 210119US00CONT
CURRENT APPLICATION NUMBER: US/09/895,072
CURRENT FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: 60/153,831
PRIOR FILING DATE: 1999-09-14
PRIOR APPLICATION NUMBER: US 09/635,872
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn Version 3.1
SEQ ID NO 2
LENGTH: 328
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-072-2

Query Match 100.0%; Score 1731; DB 9; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTFADSLRYNKKINSKFGFTSKRVPAHMPMDRIWMQELQMPFEEFDKISFKYVRIS 60
Db 1 DTFADSLRYNKKINSKFGFTSKRVPAHMPMDRIWMQELQMPFEEFDKISFKYVRIS 60
QY 61 EDQFAFSYRYVMSAVQPINISQVFEVDVTDGSLVDSREIRTLATRIHEPLSLQDPL 120
Db 61 EDQFAFSYRYVMSAVQPINISQVFEVDVTDGSLVDSREIRTLATRIHEPLSLQDPL 120
QY 121 GLEHMLINCSKMLPADITQIANNIPFQESYYDENLPPVTKSLVTNCKPVTDKIHAKYKOK 180
Db 121 GLEHMLINCSKMLPADITQIANNIPFQESYYDENLPPVTKSLVTNCKPVTDKIHAKYKOK 180

QY 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
QY 241 RPFYESMFPIPSQFELPREYRNRFLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
DB 241 RPFYESMFPIPSQFELPREYRNRFLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
QY 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328
DB 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328

RESULT 2

US-09-986-552-2
; Sequence 2, Application US/09986552
; Patent No. US20020150981A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-552-2

Query Match 100.0%; Score 1731; DB 9; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPANHPHMDIRIVMQLQDMPEEEDKTSFKRVKHS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPANHPHMDIRIVMQLQDMPEEEDKTSFKRVKHS 60
QY 61 EDMQAFASYFYILMSAVQPLNISQVDEVDTQSGVLSDRERITLATRIHELPLSLQDLT 120
DB 61 EDMQAFASYFYILMSAVQPLNISQVDEVDTQSGVLSDRERITLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHKAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHKAYKDK 180
QY 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
QY 241 RPFYESMFPIPSQFELPREYRNRFLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
DB 241 RPFYESMFPIPSQFELPREYRNRFLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
QY 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328
DB 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328

RESULT 3

US-10-023-888-5
; Sequence 5, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLUCANAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 203515US77
; CURRENT APPLICATION NUMBER: US/10/023,888

; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-888-5

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPANHPHMDIRIVMQLQDMPEEEDKTSFKRVKHS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPANHPHMDIRIVMQLQDMPEEEDKTSFKRVKHS 60
QY 61 EDMQAFASYFYILMSAVQPLNISQVDEVDTQSGVLSDRERITLATRIHELPLSLQDLT 120
DB 61 EDMQAFASYFYILMSAVQPLNISQVDEVDTQSGVLSDRERITLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHKAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHKAYKDK 180
QY 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
DB 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQLDDIRKPKRKEVCINDNDIDNHKDAQTVKAVL 240
QY 241 RPFYESMFPIPSQFELPREYRNRFLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
DB 241 RPFYESMFPIPSQFELPREYRNRFLMHMLQEMRAYRDLKRWTHCVLATLIMFTIFSFF 300
QY 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328
DB 301 AEOLIALKRRKIFPRRRIHKEASPNRIRV 328

RESULT 4

US-10-023-889-5
; Sequence 5, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBO
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-889-5

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPANHPHMDIRIVMQLQDMPEEEDKTSFKRVKHS 60
DB 1 DTFADSLRYVKNILNSKFGFTSRKVPANHPHMDIRIVMQLQDMPEEEDKTSFKRVKHS 60
QY 61 EDMQAFASYFYILMSAVQPLNISQVDEVDTQSGVLSDRERITLATRIHELPLSLQDLT 120
DB 61 EDMQAFASYFYILMSAVQPLNISQVDEVDTQSGVLSDRERITLATRIHELPLSLQDLT 120
QY 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHKAYKDK 180
DB 121 GLEHMLINCSKMLPADITQLNINIPPTQESYDYNLPVTKSLVTNCKPVTDKIHKAYKDK 180

```
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKKFCVCLNDINDHNHKAQTVKAVL 240
|
|
|
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKKFCVCLNDINDHNHKAQTVKAVL 240
|
|
|
QY 241 RDEYSEMFPIPSQFELPREYRNRFLEHMELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
|
|
|
Db 241 RDEYSEMFPIPSQFELPREYRNRFLEHMELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
|
|
|
QY 301 AEQIALKXKIFPRRRIRHKEASPNRIRV 328
|
|
|
Db 301 AEQIALKXKIFPRRRIRHKEASPNRIRV 328
|
|
|

RESULT 5
US-10-023-890-5
; Sequence 5, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYD
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023,890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-890-5
```

```
Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DTFADSLRYVKNKLNSKFGFTSRKVPAMHPMIDRIWMOELQDMPEEFDKTSFHKYRHS 60
|
|
|
Db 1 DTFADSLRYVKNKLNSKFGFTSRKVPAMHPMIDRIWMOELQDMPEEFDKTSFHKYRHS 60
|
|
|
QY 61 EDMQAFSYFYIYMSAVOPLNISQVDEVDTQSGVLSDBREIRTLATRHEHPLSLQDILT 120
|
|
|
Db 61 EDMQAFSYFYIYMSAVOPLNISQVDEVDTQSGVLSDBREIRTLATRHEHPLSLQDILT 120
|
|
|
QY 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
|
|
|
Db 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
|
|
|
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKKFCVCLNDINDHNHKAQTVKAVL 240
|
|
|
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKKFCVCLNDINDHNHKAQTVKAVL 240
|
|
|
QY 241 RDEYSEMFPIPSQFELPREYRNRFLEHMELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
|
|
|
Db 241 RDEYSEMFPIPSQFELPREYRNRFLEHMELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
|
|
|
QY 301 AEQIALKXKIFPRRRIRHKEASPNRIRV 328
|
|
|
Db 301 AEQIALKXKIFPRRRIRHKEASPNRIRV 328
|
|
|

RESULT 6
US-10-024-197-5
; Sequence 5, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS C
; FILE REFERENCE: 209794US0
; CURRENT APPLICATION NUMBER: US/10/024,197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-197-5
```

```
Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DTFADSLRYVKNKLNSKFGFTSRKVPAMHPMIDRIWMOELQDMPEEFDKTSFHKYRHS 60
|
|
|
Db 1 DTFADSLRYVKNKLNSKFGFTSRKVPAMHPMIDRIWMOELQDMPEEFDKTSFHKYRHS 60
|
|
|
QY 61 EDMQAFSYFYIYMSAVOPLNISQVDEVDTQSGVLSDBREIRTLATRHEHPLSLQDILT 120
|
|
|
Db 61 EDMQAFSYFYIYMSAVOPLNISQVDEVDTQSGVLSDBREIRTLATRHEHPLSLQDILT 120
|
|
|
QY 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
|
|
|
Db 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
|
|
|
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKKFCVCLNDINDHNHKAQTVKAVL 240
|
|
|
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKKFCVCLNDINDHNHKAQTVKAVL 240
|
|
|
QY 241 RDEYSEMFPIPSQFELPREYRNRFLEHMELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
|
|
|
Db 241 RDEYSEMFPIPSQFELPREYRNRFLEHMELOEWRAVYDKLKPWTHCVLATLIMFTISFF 300
|
|
|
QY 301 AEQIALKXKIFPRRRIRHKEASPNRIRV 328
|
|
|
Db 301 AEQIALKXKIFPRRRIRHKEASPNRIRV 328
|
|
|
```

```
RESULT 7
US-10-023-894-5
; Sequence 5, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CORNFELD, Stuart
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 5
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-894-5
```

```
Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DTFADSLRYVKNKLNSKFGFTSRKVPAMHPMIDRIWMOELQDMPEEFDKTSFHKYRHS 60
|
|
|
Db 1 DTFADSLRYVKNKLNSKFGFTSRKVPAMHPMIDRIWMOELQDMPEEFDKTSFHKYRHS 60
|
|
|
QY 61 EDMQAFSYFYIYMSAVOPLNISQVDEVDTQSGVLSDBREIRTLATRHEHPLSLQDILT 120
|
|
|
Db 61 EDMQAFSYFYIYMSAVOPLNISQVDEVDTQSGVLSDBREIRTLATRHEHPLSLQDILT 120
|
|
|
QY 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
|
|
|
Db 121 GLEHMLINCSKMLPADITQANNIPTQESYDNPPLPVTKSLVTNCKPVTDKIHKAYKDX 180
|
|
|
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPKKFCVCLNDINDHNHKAQTVKAVL 240
|
|
|
```

```
Db      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
Qy      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
Db      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
```

```
RESULT 8
US-10-306-686-2
; Sequence 2, Application US/10306686
; Publication No. US20030148460A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 230397US77DIV
; CURRENT APPLICATION NUMBER: US/10/306,686
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: 09/636,596
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-08-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-306-686-2
```

```
Query Match      100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4,3e-163;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 DTFADSLRYVNKLINSKFGFTSRKVPAHMPHMDIRIVMQLQDMFPEEFDKTSFHKVRS 60
      1 DTFADSLRYVNKLINSKFGFTSRKVPAHMPHMDIRIVMQLQDMFPEEFDKTSFHKVRS 60
Db      61 EDMQAFASYFYIYMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
      61 EDMQAFASYFYIYMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
Qy      121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTVDKIHKAYKDX 180
      121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTVDKIHKAYKDX 180
Db      121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTVDKIHKAYKDX 180
      121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTVDKIHKAYKDX 180
Qy      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
Db      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
Qy      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
Db      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
Qy      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
Db      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
```

```
RESULT 9
US-09-895-072-8
; Sequence 8, Application US/09895072
; Patent No. US200202550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119US00CONT
; CURRENT APPLICATION NUMBER: US/09/895,072
; CURRENT FILING DATE: 2001-07-02
```

```
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-895-072-8
```

```
Query Match      98.1%; Score 1698; DB 9; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 DTFADSLRYVNKLINSKFGFTSRKVPAHMPHMDIRIVMQLQDMFPEEFDKTSFHKVRS 60
      1 DTFADSLRYVNKLINSKFGFTSRKVPAHMPHMDIRIVMQLQDMFPEEFDKTSFHKVRS 60
Db      61 EDMQAFASYFYIYMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
      61 EDMQAFASYFYIYMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
Qy      121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTVDKIHKAYKDX 180
      121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTVDKIHKAYKDX 180
Db      121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTVDKIHKAYKDX 180
      121 GLEHMLINCSKMLPADITQLNINIPPTQESYVDNLPVTKSLVTNCKEPTVDKIHKAYKDX 180
Qy      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
Db      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
      181 NKRFPEIMGEELIAFMIRTNVSHVGGDDIRKPKRFVCLNDINDHNRDAQTVKAVL 240
Qy      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
Db      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
      241 RDPFESMPFIPSOQFELPREYRNRFLLMHELOEWKAYRDKLKFMTHTCVLATLIMFTISFF 300
Qy      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
Db      301 AEOLIALKKRIFPRRRIHKEASPNRIRV 328
```

```
RESULT 10
US-09-986-552-8
; Sequence 8, Application US/09986552
; Patent No. US20020150981A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-552-8
```

```
Query Match      98.1%; Score 1698; DB 9; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 DTFADSLRYVNKLINSKFGFTSRKVPAHMPHMDIRIVMQLQDMFPEEFDKTSFHKVRS 60
      1 DTFADSLRYVNKLINSKFGFTSRKVPAHMPHMDIRIVMQLQDMFPEEFDKTSFHKVRS 60
Db      61 EDMQAFASYFYIYMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
      61 EDMQAFASYFYIYMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHELPISLQDLT 120
```


Db 61 EDMQAFASYFYILMSAVQPLNISQVFEHVDITDQSGVLSDBREITLARIHDLPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNIPPTQESYDPLPVTKSLVTNCKEPTDKIHKAYXDK 180
Db 121 GLEHMLINCSKMLPANITQLNIPPTQEAAYDPLPVTKSLVTNCKEPTDKIHKAYXDK 180
Qy 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQDLDIRKNRPFVCLNDINDHNDKQACTVAVL 240
Db 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQDLDIRKNRPFVCLNDINDHNDKQACTVAVL 240
Qy 241 RDPFESMFPISQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Db 241 RDPFESMFPISQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
Db 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328

RESULT 11
US-10-023-888-10
; Sequence 10, Application US/10023888
; Publication No. US20030119086A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 203515US77
; CURRENT APPLICATION NUMBER: US/10/023, 888
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-888-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPFEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPFEEFDKTSFHKVRS 60
Qy 61 EDMQAFASYFYILMSAVQPLNISQVFEHVDITDQSGVLSDBREITLARIHDLPLSLQDLT 120
Db 61 EDMQAFASYFYILMSAVQPLNISQVFEHVDITDQSGVLSDBREITLARIHDLPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNIPPTQESYDPLPVTKSLVTNCKEPTDKIHKAYXDK 180
Db 121 GLEHMLINCSKMLPANITQLNIPPTQEAAYDPLPVTKSLVTNCKEPTDKIHKAYXDK 180
Qy 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQDLDIRKNRPFVCLNDINDHNDKQACTVAVL 240
Db 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQDLDIRKNRPFVCLNDINDHNDKQACTVAVL 240
Qy 241 RDPFESMFPISQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Db 241 RDPFESMFPISQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
Db 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328

RESULT 12
US-10-023-889-10
; Sequence 10, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOH

; TITLE OF INVENTION: DEFICIENT CELLS
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023, 889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-889-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPFEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPFEEFDKTSFHKVRS 60
Qy 61 EDMQAFASYFYILMSAVQPLNISQVFEHVDITDQSGVLSDBREITLARIHDLPLSLQDLT 120
Db 61 EDMQAFASYFYILMSAVQPLNISQVFEHVDITDQSGVLSDBREITLARIHDLPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNIPPTQESYDPLPVTKSLVTNCKEPTDKIHKAYXDK 180
Db 121 GLEHMLINCSKMLPANITQLNIPPTQEAAYDPLPVTKSLVTNCKEPTDKIHKAYXDK 180
Qy 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQDLDIRKNRPFVCLNDINDHNDKQACTVAVL 240
Db 181 NKYRFEIMGEEBEIAFKMIRTNVSHVVGQDLDIRKNRPFVCLNDINDHNDKQACTVAVL 240
Qy 241 RDPFESMFPISQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Db 241 RDPFESMFPISQFELPREYRNRFLMHLEQEMRAYRDKLKFWTHCVLATLIMFTISFF 300
Qy 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328
Db 301 AEQIALKRRKIFPRRRIRHKEASPNRIRV 328

RESULT 13
US-10-023-890-10
; Sequence 10, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOH
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023, 890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-890-10

Query Match 98.1%; Score 1698; DB 14; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPFEEFDKTSFHKVRS 60
Db 1 DTFADSLRYVKNKILNSKFGFTSRKVPAMPHMIDRIVMQELQDMPFEEFDKTSFHKVRS 60
Qy 61 EDMQAFASYFYILMSAVQPLNISQVFEHVDITDQSGVLSDBREITLARIHDLPLSLQDLT 120
Db 61 EDMQAFASYFYILMSAVQPLNISQVFEHVDITDQSGVLSDBREITLARIHDLPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLNIPPTQESYDPLPVTKSLVTNCKEPTDKIHKAYXDK 180
Db 121 GLEHMLINCSKMLPANITQLNIPPTQEAAYDPLPVTKSLVTNCKEPTDKIHKAYXDK 180

```

Db      121 GLEHMLINCSKMLPANITQLNNIPTQEAAYDPNLPVTKSLVTNCKEPTDKIHKAYDK 180
QY      181 NKYRFEIMGEEBEIAFMITNTVSHVVGQDDIRKNPKRFVCINDINDHNHDAQTAVKAVL 240
Db      181 NKYRFEIMGEEBEIAFMITNTVSHVVGQDDIRKNPKRFVCINDINDHNHDAQTAVKAVL 240
QY      241 RDFYESMFPISPQFELPREYRNRFLMHMELQEMRAYRDKLKFWTCHVLAATLIMFTIFSFF 300
Db      241 RDFYESMFPISPQFELPREYRNRFLMHMELQEMRAYRDKLKFWTCHVLAATLIMFTIFSFF 300
QY      301 AEQIILAKRKIFPFRRIHKEASPNRIRV 328
Db      301 AEQIILAKRKIFPFRRIHKEASPNRIRV 328

```

```

RESULT 14
US-10-024-197-10
; Sequence 10, Application US/10024197
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS
; FILE REFERENCE: 209794US0
; CURRENT APPLICATION NUMBER: US/10/024,197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-024-197-10

```

```

Query Match
Best Local Similarity 98.1%; Score 1698; DB 14; Length 328;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 DTFADSLRYVKNILNSKFGFTSRKVPANHPMHIDRIVMQELQDMPEEFDKTSFHKVRHS 60
Db      1 DTFADSLRYVKNILNSKFGFTSRKVPANHPMHIDRIVMQELQDMPEEFDKTSFHKVRHS 60
QY      61 EDMQFAFSFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120
Db      61 EDMQFAFSFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120
QY      121 GLEHMLINCSKMLPADITQLNNIPTQESYDYPNLPVTKSLVTNCKEPTDKIHKAYDK 180
Db      121 GLEHMLINCSKMLPADITQLNNIPTQESYDYPNLPVTKSLVTNCKEPTDKIHKAYDK 180
QY      181 NKYRFEIMGEEBEIAFMITNTVSHVVGQDDIRKNPKRFVCINDINDHNHDAQTAVKAVL 240
Db      181 NKYRFEIMGEEBEIAFMITNTVSHVVGQDDIRKNPKRFVCINDINDHNHDAQTAVKAVL 240
QY      241 RDFYESMFPISPQFELPREYRNRFLMHMELQEMRAYRDKLKFWTCHVLAATLIMFTIFSFF 300
Db      241 RDFYESMFPISPQFELPREYRNRFLMHMELQEMRAYRDKLKFWTCHVLAATLIMFTIFSFF 300
QY      301 AEQIILAKRKIFPFRRIHKEASPNRIRV 328
Db      301 AEQIILAKRKIFPFRRIHKEASPNRIRV 328

```

```

RESULT 15
US-10-023-894-10
; Sequence 10, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; APPLICANT: KORNFEID, Stuart
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; TITLE OF INVENTION: ACETYLGUCCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDS

```

```

; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-894-10

```

```

Query Match
Best Local Similarity 98.1%; Score 1698; DB 14; Length 328;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY      1 DTFADSLRYVKNILNSKFGFTSRKVPANHPMHIDRIVMQELQDMPEEFDKTSFHKVRHS 60
Db      1 DTFADSLRYVKNILNSKFGFTSRKVPANHPMHIDRIVMQELQDMPEEFDKTSFHKVRHS 60
QY      61 EDMQFAFSFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120
Db      61 EDMQFAFSFYIYMSAVQPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120
QY      121 GLEHMLINCSKMLPADITQLNNIPTQESYDYPNLPVTKSLVTNCKEPTDKIHKAYDK 180
Db      121 GLEHMLINCSKMLPADITQLNNIPTQESYDYPNLPVTKSLVTNCKEPTDKIHKAYDK 180
QY      181 NKYRFEIMGEEBEIAFMITNTVSHVVGQDDIRKNPKRFVCINDINDHNHDAQTAVKAVL 240
Db      181 NKYRFEIMGEEBEIAFMITNTVSHVVGQDDIRKNPKRFVCINDINDHNHDAQTAVKAVL 240
QY      241 RDFYESMFPISPQFELPREYRNRFLMHMELQEMRAYRDKLKFWTCHVLAATLIMFTIFSFF 300
Db      241 RDFYESMFPISPQFELPREYRNRFLMHMELQEMRAYRDKLKFWTCHVLAATLIMFTIFSFF 300
QY      301 AEQIILAKRKIFPFRRIHKEASPNRIRV 328
Db      301 AEQIILAKRKIFPFRRIHKEASPNRIRV 328

```

Search completed: July 26, 2004, 11:25:26
Job time: 26.2146 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:58 ; Search time 5.46316 Seconds
(without alignments)
3126.212 Million cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731

Sequence: 1 DTFADSLRYVKNILNSKRGF.....RKIPPRRIHKHASPNNRIIV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	124	7.2	1312	1 P1753	saccharomyc
2	106.5	6.2	971	1 OPA1	ONCNA
3	106	6.1	574	1 EZRA	STRP3
4	104	6.0	574	1 EZRA	STRP3
5	101	5.8	574	1 EZRA	STRP3
6	101	5.8	2663	1 CENE	HUMAN
7	99.5	5.7	587	1 TATR	NPVBM
8	99	5.7	385	1 GBA2	BOYBN
9	97	5.6	1961	1 MYH9	EAT
10	96	5.5	515	1 VAC1	YEAST
11	95	5.5	2867	1 RBP2	PLAVB
12	94	5.4	810	1 AMDM	YEAST
13	94	5.4	817	1 YGAB	YEAST
14	94	5.4	1978	1 MYHB	CHICK
15	93.5	5.4	353	1 STM	BORBU
16	93.5	5.4	1960	1 MYH9	HUMAN
17	91.5	5.3	587	1 VNU4	YEAST
18	91.5	5.3	610	1 VATA	TRYCO
19	91.5	5.3	1006	1 DPOL	VACCC
20	91.5	5.3	1006	1 DPOL	VACCC
21	91	5.3	582	1 TATR	NPVBM
22	91	5.3	1157	1 D3J7	HUMAN
23	90	5.2	734	1 RELA	BACSU
24	90	5.2	920	1 CAPP	STRPY
25	89.5	5.2	460	1 YS41	CAEEL
26	89.5	5.2	1005	1 DPOL	VARV
27	89.5	5.2	1701	1 MSP1	PLAFY
28	89.5	5.2	1701	1 MSP1	PLAFY
29	89	5.1	420	1 NCE2	RAT
30	89	5.1	879	1 RASO	SCHLO
31	89	5.1	1013	1 PRML	DRUME
32	89	5.1	2104	1 MYG3	SCHPO
33	88.5	5.1	463	1 POF2	SCHPO

34	88	5.1	932	1 CAPP	STRP3	Q8K873	streptococc
35	88	5.1	1118	1 CARB	YEAST	P03965	saccharomyc
36	88	5.1	2349	1 TPR	HUMAN	P12710	homo sapien
37	87.5	5.1	499	1 PPR5	CIOTE	O892X3	c biunctio
38	87.5	5.1	564	1 EZRA	STRAP	O8CWN5	staphylococ
39	87.5	5.1	661	1 HS74	TRYEB	P11145	trypanosoma
40	87.5	5.1	769	1 SWIE	KUOLA	P40418	kluyveromyc
41	87.5	5.1	1526	1 MYG2	SCHPO	Q9US16	schizosacch
42	87	5.0	372	1 Y086	METJA	Q57551	methanococc
43	87	5.0	435	1 RPOD	SORBI	Q01923	sorghum bic
44	87	5.0	439	1 FGL2	HUMAN	O14314	homo sapien
45	87	5.0	576	1 PTL	BUCBP	Q8B04	buchnera ap

ALIGNMENTS

```

RESULT 1
RA50_YEAST          STANDARD;          PRT; 1312 AA.
ID      P1753;
AC      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      DNA repair protein RAD50 (153 kDa protein).
GN      RAD50 OR YNL250W OR N0872.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=RS821;
RX      MEDLINE=9276917; PubMed=2659437;
RA      Alani E., Subbiah S., Kleckner N.;
RT      "The yeast RAD50 gene encodes a predicted 153-kD protein containing a
RT      purine nucleotide-binding domain and two large heptad-repeat
RT      regions."
RL      Genetics 122:47-57(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288c / FY1679;
RX      MEDLINE=9737992; PubMed=9234673;
RA      Sen-Gupta M., Gueldeher U., Beinhauer J.D., Fiedler T.A.,
RT      Hegemann J.H.;
RT      "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT      from the left arm of chromosome XIV from Saccharomyces cerevisiae."
RT      Yeast 13:849-860(1997).
RL      -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC      rad50/mre11 complex possesses single-strand endonuclease activity
CC      and ATP-dependent double-strand-specific exonuclease activity.
CC      Rad50 provides an ATP-dependent control of mre11 by unwinding
CC      and/or repositioning DNA ends into the mre11 active site.
CC      -1- SUBUNIT: Forms a complex with MRE11.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X14814; CA932919.1; -
CC      EMBL; X96722; CA65494.1; -
CC      EMBL; Z71526; CA96157.1; -
CC      PIR; S05808; BMEYDL.
CC      GERMOLINE; 143256; -
CC      SGD; S000519; RAD50.
CC      GO; GO:0005634; C:nucleus, IDA.
CC      GO; GO:0003690; F:double-stranded DNA binding, IDA.
CC      GO; GO:000515; F:protein binding, IPI.
CC      GO; GO:0006033; P:double-strand break repair via nonhomologous . . ; IMP.

```

DR GO:0000722; P:telomerase-independent telomere maintenance; IMP.
 DR InterPro: IPR003593; AAA_Arase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR004584; Rad50.
 DR InterPro: IPR007517; Rad50_zn_hook.
 DR Pfam: PF04423; Rad50_zn_hook; 1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRfam: TIGR00606; rad50; 1.
 DR DNA_repair; Hydrolase; ATP-binding; Coiled coil; Meiosis.
 KM NP_BIND 34 41 ATP (POTENTIAL).
 FT DOMAIN 185 347 COILED COIL (POTENTIAL).
 FT DOMAIN 403 558 COILED COIL (POTENTIAL).
 FT DOMAIN 617 722 COILED COIL (POTENTIAL).
 FT DOMAIN 734 1108 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1312 AA; 152568 MW; 58A0A173AC5677E CRC64;
 Query Match 7.2%; Score 124; DB 1; Length 112;
 Best Local Similarity 21.3%; Pred. No. 0.17; Indels 60; Gaps 14;
 Matches 68; Conservative 65; Mismatches 126; Indels 60; Gaps 14;
 QY 5 DSLRYVVKILNSKFGFTSRKVPAMPHMIDR-----IWMQLQ--DMFPSEFD--KTSF 54
 DB 189 DNLSIKKMSVDIKLKQSV-EHLKLDKDRSKAMKLNHQLQKIDQYNEVSEIESQL 247
 QY 55 HAVRSHEDMQAFSV-FYVLMASVQPLINSQVDEVDVDDQSVSDREIRTLARIEHP 113
 DB 248 NETTESKDLKFKSNODFKILSKVENLKNTKUL-----SISD-QVKLSNSIDILD 296
 QY 114 LSLQDITGLEHMLINCSKMLPADITOLNNIPPTQESYVD--PNLPVTKSLVTNCKPYTD 171
 DB 297 LSKPD---LQNLANSKVLMDKNQNLRLDTDISLMDROSLGSLNSLIRGGL-- 351
 QY 172 KIHKAYKDKKTRFELMGEELAFKIRTNVSHV---VGQLDIRKPKRFVC--LMDN 225
 DB 352 EAGKETEKRNHLSLKE--AFQHKFQGLSNINSMDAQVNHMSQKAFISQDLTDT 408
 QY 226 IDHNHKAQTVAAVLDPEFSWMPPIPSQFELPREYRNFLE----- 266
 DB 409 IDQFAKDITLKETNSLDLKSITVDSQNLFNKPKRSKLIHDEELAKKSPKSLSTQD 468
 QY 267 --MHLEQWRAVPDKLKFV 283
 DB 469 SLNHELENKTYEKELQSW 487
 RESULT 2
 OPAL_ONCMA STANDARD; PRT; 971 AA.
 ID OPAL_ONCMA 093248;
 AC 093248;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dymamin-like 120 kDa protein, mitochondrial precursor (120-kDa G
 protein expressed in motor neurons) (mg120).
 GN OPAL.
 OS Oncorhynchus masou (cherry salmon) (Masu salmon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Procacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OX NCBI_TaxID=8020;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98372640; PubMed=9708909;
 RA Kubokawa K., Miyashita T., Kubo Y.;
 RT "Isolation of a cDNA for a novel 120-kDa GTP-binding protein expressed
 in motor neurons in the salmon brain.";
 RL FEBS Lett. 431:231-235(1998).
 CC 1- FUNCTION: May be involved in mitochondrial biogenesis.
 CC 1- SUBCELLULAR LOCATION: Mitochondrion (by similarity).
 CC 1- TISSUE SPECIFICITY: Strongly expressed in the brain, ovary and
 CC skeletal muscle. In the brain, expression of the mRNA was observed
 CC specifically in motor neurons, in nucleus oculomotorius, in
 CC nucleus valvulae lateralis, in the medulla oblongata and in the

CC 1- Spinal cord.
 CC 1- SIMILARITY: Belongs to the dymamin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
 CC or send an email to license@isb-stb.ch).
 DR EMBL: AB012720; BA032279.1; -.
 DR PIR: T00394; T00394.
 DR GO: GO:0030425; C:cytosol; ISS.
 DR GO: GO:0030061; C:mitochondrial cristae; ISS.
 DR GO: GO:0005758; C:mitochondrial intermembrane space; ISS.
 DR GO: GO:0005741; C:mitochondrial outer membrane; ISS.
 DR GO: GO:0003924; F:GTPase activity; IDA.
 DR GO: GO:0000287; F:magnesium ion binding; ISS.
 DR GO: GO:0019896; P:axon transport of mitochondrion; ISS.
 DR GO: GO:0007007; P:inner mitochondrial membrane organization a. .; ISS.
 DR GO: GO:0000266; P:mitochondrial fission; ISS.
 DR GO: GO:0008053; P:mitochondrial fusion; ISS.
 DR GO: GO:0045768; P:positive regulation of anti-apoptosis; ISS.
 DR GO: GO:0007601; P:division; ISS.
 DR InterPro: IPR001401; Dymamin.
 DR Pfam: PF00350; dymamin; 1.
 DR SMART: SM00195; DYNAMIN.
 DR SMART: SM00053; DYNC; 1.
 KM Motor protein; GTP-binding; Mitochondrion; Transit peptide.
 FT TRANSIT 1 971 MITOCHONDRION (POTENTIAL).
 FT NP_BIND 307 314 DYNAMIN-LIKE 120 KDA PROTEIN.
 FT NP_BIND 410 414 GTP (POTENTIAL).
 FT NP_BIND 478 481 GTP (POTENTIAL).
 SQ SEQUENCE 971 AA; 112069 MW; 462958C03C1A4F02 CRC64;
 Query Match 6.2%; Score 106.5; DB 1; Length 971;
 Best Local Similarity 23.0%; Pred. No. 2.4; Indels 85; Gaps 14;
 Matches 62; Conservative 43; Mismatches 79; Indels 85; Gaps 14;
 QY 7 LRYVVKILNSKFGFTSRKVPAMPHMIDRIVMQLQDMFPSEFDKTSFKHRSSEMQ 64
 DB 102 LRYI-VLSAVAGGYTAKK-----TYEWMKMLPD----- 130
 QY 65 FARSFYVYLSA---VQPLNISQVPEVDVDDQ-SGVLSSREIRTLARIEHPISQDIT 120
 DB 131 -----MSANNWIPDFWELSDQIDLDKLTKLTP--ELSEIAKLPELP---DF- 174
 QY 121 GLEHMLINCSKMLPADITOLNNIPPTQESYVDNLPVTKSLVTNCKPYTDKIHKAYKDX 180
 DB 175 -----DKIGENFTFLKSLIFPAEAPGPTPVAAATEAVTATPEASDQKPKSSDX 224
 QY 181 NKTRFELMGEELAFKIRTNVSH--VGQLDIRKPKRFVC--LMDN 225
 DB 225 EKV-----DQQLIELRTQMKYQEMERLEKENDKLVQLQDEKGIHQK---IK 273
 QY 238 AVLRDPEFSWMPPIPSQF-----LPR 258
 DB 274 KSLIDMYSEVLDLSDPSDENVYTDHLPR 302
 RESULT 3
 EZRA_STRPY STANDARD; PRT; 574 AA.
 ID EZRA_STRPY 09A059;
 AC 09A059;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Septation ring formation regulator ezra.
 GN EZRA OR SPY0728.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

```
CC Streptococcus.
OK NCBI_TaxId=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -!- FUNCTION: Negative regulator of ftsZ ring formation; modulates the
CC frequency and position of ftsZ ring formation. Inhibits ftsZ ring
CC formation at polar sites. Interacts either with ftsZ or with one
CC of its binding partners to promote depolymerization (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Membrane-associated. Colocalized with ftsZ to
CC the nascent septal site (By similarity).
CC -!- SIMILARITY: Belongs to the ezra family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB06524; AK33678.1; -.
DR HAMAP; MF_00728; -.
KW Cell division; Septation; Transmembrane; Coiled coil;
KW Complete proteome.
FT DOMAIN 1 7 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 8 26 POTENTIAL.
FT DOMAIN 27 574 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 102 148 Coiled coil (Potential).
FT DOMAIN 276 366 Coiled coil (Potential).
SQ SEQUENCE 574 AA; 66120 MW; 91FF87736186827 CRC64;

Query Match 6.1%; Score 106; DB 1; Length 574;
Best Local Similarity 21.8%; Pred. No. 1.3; Indels 58; Gaps 12;
Matches 61; Conservative 48; Mismatches 113;

QY 6 SLRYVVKILNSKFGTSRKVPAMPMIDRIYVQELQDMFPE---EFDKTSFHKVRHSD 62
DB 126 SIREALNIKEQ---EEKNSARVTHALD--LYEKLOASISENEDFGSTMPEDIKQMK 179
QY 63 MOPAFSYFYILMSAVOPLNISQVFD--EVDTDQSGVLSDRRIATRIHELPISLODL- 119
DB 180 IETEFISQFVALNSGDPVEASEVLDRAEHTIALGQITQIPAIYAKLEDDEFPQDLLE 239
QY 120 TGLEHMLINCSKMLPADITQNLNIPPTQESYDP-----NLPTVTKSLVTKCKP-VTDK 172
DB 240 TGYRRL-----EENHYHPEKNIEARFOETRESIRANSSELVTLDD 279
QY 173 IHKAYKDKKKYR-----PEIMGEEELIAFKMIRTVSHVQQLDIRKPKFVCLNDNT 226
DB 280 IDRAEENHTIQRIDSLYEVEEREIIAAYKAAKNSKMLPYLEHVKNNQ---LKDEI 336
QY 227 DH-----NKKDAQTVAKVLRDPE---SMPEIPISQEL 256
DB 337 AELSKYILSETESLTVKAFKDIKEIDEDSTLAVERGL 376

RESULT 4
EZRA_STRP3 STANDARD; PRT; 574 AA.
AC 08K39;
DT 15-MAR-2004 (rel. 43, Created)
DT 15-MAR-2004 (rel. 43, Last sequence update)
DT 15-MAR-2004 (rel. 43, Last annotation update)
DE Septation ring formation regulator ezra.
```

```
GN EZRA OR SPVW3_0477 OR SPB1377.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobactillales; Streptococcaceae;
CC Streptococcus.
OK NCBI_TaxId=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Bares S.B., Sylva G.L., Barbhan K.D., Lei B., Hoff J.S.,
RA Mamarrella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055(2003).
CC -!- FUNCTION: Negative regulator of ftsZ ring formation; modulates the
CC frequency and position of ftsZ ring formation. Inhibits ftsZ ring
CC formation at polar sites. Interacts either with ftsZ or with one
CC of its binding partners to promote depolymerization (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Membrane-associated. Colocalized with ftsZ to
CC the nascent septal site (By similarity).
CC -!- SIMILARITY: Belongs to the ezra family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE014146; AAU79084.1; -.
DR HAMAP; MF_00728; -.
DR HAMAP; MF_00728; -.
KW Cell division; Septation; Transmembrane; Coiled coil;
KW Complete proteome.
FT DOMAIN 1 7 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 8 26 POTENTIAL.
FT DOMAIN 27 574 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 102 148 Coiled coil (Potential).
FT DOMAIN 276 366 Coiled coil (Potential).
SQ SEQUENCE 574 AA; 66062 MW; 25F877273686825 CRC64;

Query Match 6.0%; Score 104; DB 1; Length 574;
Best Local Similarity 22.5%; Pred. No. 1.9; Indels 66; Gaps 14;
Matches 64; Conservative 48; Mismatches 106;

QY 6 SLRYVVKILNSKFGTSRKVPAMPMIDRIYVQELQDMFPE---EFDKTSFHKVRHSD 62
DB 126 SIREALNIKEQ---EEKNSARVTHALD--LYEKLOASISENEDFGSTMPEDIKQMK 179
QY 63 MOPAFSYFYILMSAVOPLNISQVFD--EVDTDQSGVLSDRRIATRIHELPISLODL- 119
DB 180 IETEFISQFVALNSGDPVEASEVLDRAEHTIALGQITQIPAIYAKLEDDEFPQDLLE 239
QY 120 TGLEHMLINCSKMLPADITQNLNIPPTQESYDP-----NLPTVTKSLVTKCKP-VTDK 172
DB 240 TGYRRL-----EENHYHPEKNIEARFOETRESIRANSSELVTLDD 279
```

```

QY 173 IHKAYDKKKYR-----FEINGEEIAFKMIRTN-----VSHV---COL-DDIRKN 215
DB 280 LDRAREENTHIERIDSLYEVEFEREIAKVAAKSKMLPRILAVKRNNEQLDEIARL 339
QY 216 PKKVCGLNDNIDHNHKAQTKAVLRDPYE---SMFPISQPEL 256
DB 340 SRKYL-----SETESLTVKAFKDIKEIDSTLVAEQGL 376

RESULT 5
EZRA_STRPE
ID EZRA_STRPE8 STANDARD; PRT; 574 AA.
AC Q8P1M3;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Separation ring formation regulator ezra.
GN EZRA OR SPYM18.0796.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.U., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.
RT Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
CC -1- FUNCTION: Negative regulator of ftsZ ring formation; modulates the
CC frequency and position of ftsZ ring formation. Inhibits ftsZ ring
CC formation at polar sites. Interacts either with ftsZ or with one
CC of its binding partners to promote depolymerization (by
CC similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Colocalized with ftsZ to
CC the nascent septal site (By similarity).
CC -1- SIMILARITY: Belongs to the ezra family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE010010; AAL97460.1; -.
DR HAMAP; MF_00728; -.
KW Cell division; Septation; Transmembrane; Coiled coil;
KW Complete proteome.
FT DOMAIN 1 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 7 POTENTIAL.
FT DOMAIN 8 26 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 27 574 COILED COIL (POTENTIAL).
FT DOMAIN 102 148 COILED COIL (POTENTIAL).
FT DOMAIN 276 366 COILED COIL (POTENTIAL).
SQ SEQUENCE 574 AA; 66061 MW; 76A95C1FC4EB0EFC CRC64;

Query Match 5.8%; Score 101; DB 1; Length 574;
Best Local Similarity 21.1%; Pred. No. 3.2;
Matches 60; Conservative 48; Mismatches 110; Indels 66; Gaps 12;
DB 6 SARYNKILNSKFGFISRVPAHMHMDRIYMOGLQDMFPE--EFPKTFHAKYRSHED 62
DB 126 SREALNLTKEQ-----EENKSAKVTHAD--LYEKLOASISENEDNFGSTYPEIDKQKN 179
QY 63 MFAFSYFYUWSAVOPLNISQVFD--EVDTPQSGVLDRERILATRIHILPLSLQDL- 119
DB 180 IETPSQFVALNSSQDPVASEVLDRABEHTIALGQITRQPAIYAKKEDDFPPQDLBLE 239

```

```

QY 120 TGEHMLINCSKMLPADITQNLNPPQOSYDP-----NLPPVTKSLVTNCKP-VTDK 172
DB 240 TGYRRL-----EENYHPEKNIEARQOELRESIRASSELVITLD 279
QY 173 IHKAYDKKKYR-----FEINGEEIAFKMIRTNVSHVQGLDIRKN----- 215
DB 280 LDRAREENTHIERIDSLYEVEFEREIAKVAAKSKMLPRILAVKRNNEQLKNEIARL 339
QY 216 PKKVCGLNDNIDHNHKAQTKAVLRDPYE---SMFPISQPEL 256
DB 340 SRKYL-----SETESLTVKAFKDIKEIDSTLVAEQGL 376

RESULT 6
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis."
RT Nature 359:536-539(1992).
RL ENBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1."
RL J. Cell Biol. 143:49-63(1998).
RN [4]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RT microtubules."
RL J. Biol. Chem. 275:30451-30457(2000).
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE. AND IS
CC QUANTITATIVELY REDUCED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARITY: Belongs to the kinesin-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

CC or send an email to license@sib-sib.ch.
 DR EMBL; Z15005; CAA78727.1; -
 DR PIR; S28261; S28261.
 DR HSSP; P17119; 3KAR.
 DR GeneW; HGNC:1656; CENPE.
 DR GK; Q02224; -
 DR MIM; 117143; -
 DR GO; GO:0005699; C.kinetochore; TAS.
 DR GO; GO:0005634; C.nucleus; TAS.
 DR GO; GO:0008350; F.kinetochore motor activity; TAS.
 DR GO; GO:0000067; P.DNA replication and chromosome cycle; TAS.
 DR GO; GO:0007079; P.mitotic chromosome movement; TAS.
 DR GO; GO:0007080; P.mitotic metaphase plate congression; TAS.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS00467; KINESIN_MOTOR_DOMAIN2; 1.
 DR Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis.
 KM Cell cycle; Centromere; Lipoprotein; Premylation.
 FT DOMAIN 1 335 KINESIN-MOTOR.
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT NP BIND 2472 2663 GLOBULAR (POTENTIAL).
 FT NP BIND 86 93 APP (BY SIMILARITY).
 FT LIPID 2660 2660 S-farnesyl cysteine.
 SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8C8 CR64;
 Query Match 5.8%; Score 101; DB 1; Length 2663;
 Best Local Similarity 20.5%; Pred. No. 21;
 Matches 60; Conservative 47; Mismatches 127; Indels 58; Gaps 10;
 QY 33 IIRIMQEIQDMFPPEEEDTSPHKVRHSEDMQFAFSYFYIYMSAVQPLNISQVFEVDUTD 92
 DB 1972 IELQKREIQQLRVKEDVNMHSHKINEMQQLKQEPYLCQCEMDNFQLRKQHE--- 2027
 QY 93 QSGVLSDEIRTLARIRHPLSLQDLGLEMILNCSKMLPADITQNLNIPQOESYYD 152
 DB 2028 ---SLEIRIVAEKREDLRIKESLKNERQDFATLRKEMTARQNHQKPKERLLSD 2082
 QY 153 ENLPVYTSGLVNTGCPVTDKIKHAYKDKNYK---RPEINGEEELAFKMTIRTVSHVG 207
 DB 2083 -GQOHLMESLRKESCRKIKELT-KRYSEMDHYECNLRLSLDEKIEFHRIMKKLKLYLS 2140
 QY 208 QLDDIRKXNRKRFVCLND-----NIDNNHDAQ----- 235
 DB 2141 VYTKIKEQHE--CLNKREMDIDVEYKQKELLIKIQHLOQDCDVPSELRDLKLNQMD 2198
 QY 236 --VKAFLRDVYSNFP-IPSGELPREYRNRLHNLHLEQEMRAYR---DKLK 281
 DB 2199 LHIEELIKDFSESEFPSTKTEFQ--QVLSNRKEMTQGLEMLNTRFDIEKLK 2248
 RESULT 7
 TATR NPVEM STANDARD; PRT; 587 AA.
 AC P33245;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Trans-activating transcriptional regulatory protein (Immediate early protein 1) (IE-1).
 GN IE1.
 OS Bombyx mori nuclear polyhedrosis virus (BmNPV).
 CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 CC Nucleopolyhedrovirus.
 CC NCBI_Taxid=10458;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92162753; PubMed=1536885;
 RA Huybrechts R., Guarino L., van Brusseel M., Vuleteke V.;
 RT "Nucleotide sequence of a transactivating Bombyx mori nuclear

RT polyhedrosis virus immediate early gene";
 RL Biochim. Biophys. Acta 1129:328-330(1992).
 CC -1- FUNCTION: Regulatory transcriptional protein, which trans-activates gene expression from early baculovirus promoters. Can also trans-activate its own promoter, suggesting that it is autoregulated during normal infection of insect cells.
 CC -1- SIMILARITY: TO OTHER BACULOVIRUSES IE-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X58442; CAA4348.1; -
 DR PIR; S20596; S20596.
 DR InterPro; IPR005092; TATR.
 DR Pfam; PF03430; TATR; 1.
 KW Early protein; Transcription regulation; Trans-acting factor.
 SQ SEQUENCE 587 AA; 67495 MW; 5896B0C9625ACB59 CR64;
 Query Match 5.7%; Score 99.5; DB 1; Length 587;
 Best Local Similarity 23.4%; Pred. No. 4.2; Indels 53; Gaps 11;
 Matches 58; Conservative 34; Mismatches 103; Indels 53; Gaps 11;
 QY 39 QELQDMFPPEE-----FDKTSFHKVRHSEDM-QFAFSYFYIYMSAVQPLNI-----SQ 84
 DB 189 QEITHYFTNDFAPYLRFPDNDYNSRFSQDMSETGYMEVVKSEVKEPEILFAKYVSV 248
 QY 85 VPDEVLDQSGVLSDEIRTLARIRH-----ELPISLDLGLGEMILNCSK 131
 DB 249 VVVEY--TNMYWMDNRFPVVTFDKIRPMISYNLKETGELIRPS--QDVCDDEFAQNCCK 306
 QY 132 MLPADITQNLNIPPTQESYYDPMIPVTKSLVT-----NCKPVTDKIKHAYKDKNYK 183
 DB 307 CHEVDVA--HTFKAALTSYFNLDMYAQTFTVTLQSLGRKCGFLSLKLYEYQDKNLF 364
 QY 184 RPEIM-----GEEELAFK--MRTNVSHVVGQLDIRK-----NPKRFVCLNDINRN 229
 DB 365 TLPIMSRKSNIEFTASNNFVSPYVSQLKXSEIRKVKPDPNPPKXYVDNLILVN 424
 QY 230 HKDAQTYK 237
 DB 425 KKSTLTLYK 432
 RESULT 8
 GBA2 SOYBN STANDARD; PRT; 385 AA.
 ID GBA2 SOYBN
 AC P91163;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Guanine nucleotide-binding protein alpha-2 subunit (GP-alpha-2).
 GN GBA2 OR GA2.
 OS Glycine max (Soybean).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 CC Euphorbia; Fabales; Fabaceae; Papilionoidae; Phaseoleae; Glycyne.
 CC NCBI_Taxid=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Corsoy;
 RX MEDLINE=97156166; PubMed=9002626;
 RA Gotor C., Lam E., Cejudo F.J., Romero L.C.;
 RT "Isolation and analysis of the soybean GBA2 gene (cDNA), encoding a new member of the plant G-protein family of signal transducers";
 RL Plant Mol. Biol. 32:1227-1234(1996).
 CC -1- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are involved as modulators or transducers in various transmembrane signaling systems.

```

CC -1- SUBUNIT: G proteins are composed of 3 units; alpha, beta and
CC gamma. The alpha chain contains the guanine nucleotide binding
CC site.
CC -1- SIMILARITY: Belongs to the G-alpha family. Subfamily 2 (G(s)).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X55882; CAA64934.1; -.
CC HSPB: P10824; IAS3.
CC InterPro: IPR001019; Gprotein_alpha.
CC Pfam: PF00503; G-alpha; 1.
CC PRINTS: PR00318; GPROTEIN.
CC ProDom: PD000281; Gprotein_alpha; 1.
CC SMART: SM00275; G_alpha; 1.
CC GTP-binding; Transducer; ADP-ribosylation.
CC NP_BIND 47 54 GTP (BY SIMILARITY).
CC FT NP_BIND 220 224 GTP (BY SIMILARITY).
CC FT NP_BIND 289 292 GTP (BY SIMILARITY).
CC FT MOD_RES 192 192 ADP-RIBOSYL[1] (BY ACTION OF CTX).
CC SEQUENCE 385 AA; 45096 MW; 7721B1E307067075 CRC64;

Query Match 5.7%; Score 99; DB 1; Length 385;
Best Local Similarity 22.6%; Pred. No. 2.7;
Matches 57; Conservative 44; Mismatches 101; Indels 50; Gaps 11;

QY 37 VNOEIQDMPEEFEDDT--SPKRVHSEDMQAFSYFYLYMSAVOPLNI---GOVPEV 89
DB 56 IKQIKLFGTGFDEALSKYLPVHA-----NVYQITLHDSKEPQN 101
QY 90 DFDQSGVLSDEIRFLAIRHELPLSLQDLGLEHMLINCSKMLPADITQLNIPQES 149
DB 102 DVDSKRYVINSNKKIGELILEIGRL-DYPL-----SKELAQEIENMKDPAIOET 153
QY 150 Y---YDPMLEPVTKSLVNTCKEVDTKIKHAYKDKKKY-RPEIMGEELAFKMIPTN-VSH 204
DB 154 YARGSELDQIPDCTDYFMENTLQRLSDANVPYTEDVLYARVTTGYVEIQSFVGENKSD 213
QY 205 VVGQLDDI--RKNPRKFCVCLNDNIDHNKDAQTVKAVLRDPFESMPPIPOFELPREYR 261
DB 214 EYVRLPFDVGGQNRERRKMIHLPEGV---SAVIFCAIASEVDQLF-----EDENR 260
QY 262 NRELIHHEIQEW 273
DB 261 NRMETKELFEW 272

RESULT 9
MYH9_RAT STANDARD; PRT; 1961 AA.
AC Q62812;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
CC cell shape, and specialized functions such as secretion and
CC capping.

```

```

CC -1- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
CC regulatory light chain subunits (MLC-2).
CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U1463; AAA74950.1; -.
CC HSPB: P10587; 1BR2.
CC InterPro: IPR000048; IQ_region.
CC InterPro: IPR001609; myosin_head.
CC InterPro: IPR004009; Myosin_N.
CC InterPro: IPR002928; Myosin_tail.
CC Pfam: PF00612; IQ; 1.
CC Pfam: PF00663; myosin_head; 1.
CC Pfam: PF02736; Myosin_N; 1.
CC Pfam: PF01576; Myosin_tail; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00015; IQ; 1.
CC SMART: SM00242; MYSC; 1.
CC PROSITE: PS50096; IQ; 1.
CC KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
CC Coiled coil; Multigene family.
CC DOMAIN 1 778 MYOSIN HEAD-LIKE.
CC FT DOMAIN 779 808 IQ.
CC FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
CC FT NP_BIND 174 181 ATP (POTENTIAL).
CC FT DOMAIN 654 676 ACTIN-BINDING.
CC SEQUENCE 1961 AA; 226336 MW; 9B9876D9681FB19E CRC64;

Query Match 5.6%; Score 97; DB 1; Length 1961;
Best Local Similarity 28.2%; Pred. No. 29;
Matches 35; Conservative 26; Mismatches 51; Indels 12; Gaps 5;

QY 102 IRT-LATRIHELPLSLQDLGLEHMLINCSKMLPADITQL-NNIPPOGSSYYDPLPVY 159
DB 1267 VRTLELDKYSKIQVELDSTGLNQSXSXTDFSLBSQLODTGLQENRQKL- 1325
QY 160 KSLVNTCKEVDTKIKHAYKDKKKYRPEIMGEELAFKMIPTNVSIVGQLDIRKNPKYF 219
DB 1326 -SLSTYLKQMED-----EKNSFREOLEEERBEAKRNLEKQIATLHAQVTDKMKMEDG 1377
QY 220 V-CL 222
DB 1378 VGCL 1381

RESULT 10
VAC1_YEAST STANDARD; PRT; 515 AA.
AC P32609;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE VAC1 protein.
GN VAC1 OR PEP7 OR VPS19 OR YDR323C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

```



```
CC STRAIN=LW148;
RX MEDLINE=92112720; PubMed=1730622;
RA Weisman J.S., Wickner W.;
RT "Molecular characterization of VAC1, a gene required for vacuole
RT inheritance and vacuole protein sorting.";
RL J. Biol. Chem. 267:618-623(1992).
CC -1- FUNCTION: REQUIRED FOR VACUOLE SEGREGATION AND VACUOLE PROTEIN
CC SORTING. POSSIBLY PART OF A COMPLEX WHICH TETHERS THE VACUOLE
CC MEMBRANE TO MICROTUBULES, EITHER DIRECTLY OR VIA KINISIN OR
CC DYMEIN-LIKE MOTOR PROTEINS. PROBABLY FUNCTIONS IN SEVERAL
CC INTERORGANELLE TRAFFIC PATHWAYS.
CC -1- SUBCELLULAR LOCATION: PERIPHERALLY BOUND ELEMENT OF THE
CC CYTOSKELETON OR VACUOLE (POTENTIAL).
CC -1- SIMILARITY: contains 1 C2H2-type zinc finger.
CC -1- SIMILARITY: contains 2 FYVE-type zinc fingers.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-109 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80596; AAA35203.1; -
DR GEMonline; 140815; -
DR SGD; S0002731; PEP7.
DR GO; GO:0005737; Cytoplasm; IDA.
DR GO; GO:0019897; Cytoplasmic to plasma membrane; IDA.
DR GO; GO:0006896; P:Golgi to vacuole transport; IGI.
DR GO; GO:0006904; P:nonspecific vesicle docking; IMP.
DR GO; GO:0006906; P:nonspecific vesicle fusion; IMP.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF00096; zf_C2H2; 1.
DR SMART; SM00064; FYVE; 2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS01078; ZF_FYVE; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR Zinc-finger; Repeat.
DR ZNF_FING 6 29 C2H2-TYPE.
FT ZN_FING 72 137 FYVE-TYPE 1.
FT ZN_FING 215 297 FYVE-TYPE 2.
SQ SEQUENCE 515 AA; 59469 MW; 02C204E1BCA8CACC CRC64;

Query Match 5.5%; Score 96; DB 1; Length 515;
Best Local Similarity 22.4%; Pred. No. 6;
Matches 57; Conservative 39; Mismatches 96; Indels 62; Gaps 12;

QY 28 HRP-----HMIDRIWMQELQDMFPEEPDKTSFKVHSHSEDMQAFSYFYLYMSAVQPL 80
DB 281 HPIPIRLCSHCIDM-----FIGKFKK-----DKRMPLSGITAKTDSNQ-- 321
QY 81 NISQVFEVDYDQSGVLSDRSIRLTATRIHEL-PLSLDTLGLBMLINCSKMLPADITQ 139
DB 322 NISKYIDSLPIFEDSLNKLKAVETAKDSENTLDPRGNLMDLARKRYKLINSFNLVNTLTRQ 381
QY 140 L-----NNIPFOESYVDNLPVYTSVLTNCKCPYDKTHKXKDKKXKRF 185
DB 382 LLSVBPQSHLERQONSLKIASAAYINKEIPL-KSLPAIINP--EGHKNEGQKXEP 437
QY 186 EI-----MGEERIAK-----MIRTNVSHV-GQLDDIRKNFR--KEVCIDNDIDH 228
DB 438 EYKXLSQMLTENLTLTKVXELBEELMWLKEGYSLESTIOPYKKQRRLREIVTLTKNLE 497
QY 229 HKKDACTYKAVLRD 242
DB 498 LHSRIHTVQSKLD 511
```

```
RESULT 11
RBP2_PLAVB STANDARD; PRT; 2867 AA.
AC 000759; Q9N2M3;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Reticulocyte binding protein 2 precursor (PVRBP-2).
GN RBP-2 OR RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=1273;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
RX MEDLINE=20299192; PubMed=10838229;
RA Galinski M.R., Xu M., Barnwell J.W.;
RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares
RT structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa
RT thoptry protein family.";
RT Mol. Biochem. Parasitol. 108:257-262(2000).
RN [2]
RP SEQUENCE OF 1189-2439 FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RT Cell 69:1213-1226(1992).
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
CC human reticulocyte cells.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF184623; AAF76525.1; -
DR HSSP; P03069; ICGM; Signal; Transmembrane; Repeat.
DR KMWalaria; Receptor; Signal; Transmembrane; Repeat.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.
FT DOMAIN 22 2805 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2806 2826 POTENTIAL.
FT DOMAIN 2827 2867 CYTOSOLASMIC (POTENTIAL).
FT DOMAIN 44 133 ASN-RICH.
FT DOMAIN 560 758 LYS-RICH.
FT DOMAIN 1112 1285 LYS-RICH.
FT DOMAIN 2758 2785 7 X 4 AA TANDEM REPEATS OF H-D-D-T.
FT REPEAT 2758 2761 1.
FT REPEAT 2762 2765 2.
FT REPEAT 2766 2769 3.
FT REPEAT 2770 2773 4.
FT REPEAT 2774 2777 5.
FT REPEAT 2778 2781 6.
FT REPEAT 2782 2785 7.
SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;

Query Match 5.5%; Score 95; DB 1; Length 2867;
Best Local Similarity 20.5%; Pred. No. 66;
Matches 56; Conservative 55; Mismatches 78; Indels 84; Gaps 14;

QY 7 LRYVKK-----INSKFGFTSRKVPAPMPHMDIRIWMQELQDMFPEEPDKTSFKVYR 59
DB 2193 LAYIKKNYEDTVQDVLTINLHFNTRKQVSNH-----PNNFDKSN-----K 2232
QY 59 HSEDMQAFSYFYLYMSAVQPLNISQVFD-EVDYDQSGVLSDRSIRLTATRIHEL-PLSL 116
DB 2233 SSEELTKAVYDSKTIISKLKGV-IIEVNTEKNTIES--SAKETLEALYNELKNNKTSL 2288
QY 117 QDL-----TGLEHMLINCSKMLPADITQ-LNNIPFOESYVDNLPVYTSVLTNCKCPV 169
```


DR EMBL; X82775; CAA58017.1; -
 DR EMBL; Z72983; CAA87225.1; -
 DR FIR; S53921; S53921.
 DR GenOnline; 141510; -
 DR SGD; S0003430; YGR198M.
 DR InterPro; IPR008941; TPR-1like.
 KW Hypothetical protein.
 SO SEQUENCE 817 AA; 95365 MW; AC42730CB93CE4B CRC64;

Query Match 5.4%; Score 94; DB 1; Length 817;
 Best Local Similarity 20.2%; Pred. No. 17;
 Matches 73; Conservative 44; Mismatches 121; Indels 124; Gaps 14;

1 DPFADSLRVNKLINSKFGFTSRKYPALHPHMIDRIY-----MDEL 41
 260 EINDALAEHNVFLQYRFRTTHKIKXNPSWYNITISMEXTPOSIEVSKTAMFLPQNL 319
 42 QDMPEEPDKTSF-----HKVHSEDMQAFSFFYILMSAVQPL 80
 320 SGNNDDEIKKTKFKESILNFVFNKYNDKYQLHDNHRDIIISFIDAYSFILQNSKTD 379
 81 NISQVDEVDIDQSVLSDRREIRLATRI-----HELPLSIQ--DLTGLEHML----- 127
 380 SIENVFD-----YDNTVSTRATSLNSFKRYNPLMSQSGLMLNSTRCVYPG 429
 128 NCSKMLPADITQLNINIPPTQSESYDPNLPVTKSLVTNCKEYTDKIKHAYDKKKYRPEI 187
 430 NISKVLTNAMSTLYEIRKQDLFVSN--NLTSYLCNAMMLSTK---EKDNADVEEQ 482
 188 MEESEIAFMKMITNVSHVVGOLDIR-----KNPR----- 217
 483 EEESEKALREILOFKKSYTLAQORHLETAIKTLESILSKNPPYKAMHLALCRSVQEDK 542
 218 ---KFVC---LNDNIDHN-----HKDQTVKAVLRDFY---ESMEPIPSQF 254
 543 EMSYKIVCSVLNAMESLQNTLLNDKQFIHLTL-QQLALIEIFGLTALETLPVVF 601
 255 EL 256
 602 EL 603

RESULT 14
 MYHB_CHICK STANDARD; PRT; 1978 AA.
 AC P10587;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, gizzard smooth muscle.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88118918; PubMed=2892941;
 RA Yanagisawa M., Hamada Y., Katsuragawa Y., Imamura M., Mikawa T.,
 RA Masaki T.;
 RT "Complete primary structure of vertebrate smooth muscle myosin heavy
 RT chain deduced from its complementary DNA sequence. Implications on
 RT topography and function of myosin.";
 RL J. Mol. Biol. 1981;143-157(1987).
 RN [2]
 RP REVISIONS.
 RA Masaki T.;
 RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RP SEQUENCE OF 1-203.
 RX MEDLINE=88032919; PubMed=3312184;
 RA Maizumi T., Onishi H., Yajima E., Matsuda G.;
 RT "Amino acid sequence of the amino-terminal 24 kDa fragment of the

RT heavy chain of chicken gizzard myosin.";
 RL J. Biochem. 102:133-145(1987).
 RL [4]
 RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-818, AND REVISIONS 204-215.
 RX MEDLINE=98442652; PubMed=9741621;
 RA Dominguez R., Freyzon Y., Tryba K.M., Cohen C.;
 RT "Crystal structure of a vertebrate smooth muscle myosin motor domain
 RT and its complex with the essential light chain: visualization of the
 RT pre-power stroke state";
 RL Cell 94:559-571(1998).
 CC -1- FUNCTION: Muscle contraction.
 CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 CC and 2 regulatory light chain subunits of the myofibrils.
 CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 CC meromyosin (MM) and 1 heavy meromyosin (HMM). It can later be
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped
 CC subfragment (S2).
 CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -1- SIMILARITY: Contains 1 IQ domain.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X06546; CAA29793.1; -
 DR FIR; S03166; S03166.
 DR PDB; 1BR1; 09-SEP-98.
 DR PDB; 1BR2; 09-SEP-98.
 DR PDB; 1BR4; 09-SEP-98.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR001609; myosin head.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRODOM; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Calmodulin-binding; Multigene family;
 KW 3D-structure.
 FT INT MET 0 0
 FT DOMAIN 1 790
 FT DOMAIN 791 820
 FT DOMAIN 849 1978
 FT DOMAIN 1978 1978
 FT NP BIND 176 183
 FT DOMAIN 666 688
 FT DOMAIN 767 781
 FT MOD RES 1 1
 FT MOD RES 127 127
 FT MOD RES 127 127
 FT CONFLICT 127 127
 FT CONFLICT 204 215
 FT STRAND 34 37
 FT STRAND 43 47
 FT STRAND 56 60
 FT STRAND 66 69
 FT TURN 72 73
 FT STRAND 75 76
 FT STRAND 76 76

MYOSIN HEAD-LIKE.
 IQ.
 RODLIKE TAIL (S2 AND IMM DOMAINS).
 COILED COIL (POTENTIAL).
 ATP.
 ACTIN-BINDING.
 ACTIN-BINDING.
 BLOCKED.
 METHYLATION (TR1-) (POTENTIAL).
 MISSING (IN REF. 3).
 KDTSIQGPSFS -> RTASLKVHLP (IN REF. 1).

```
FT HELIX 80 82
FT TURN 83 84
FT STRAND 87 87
FT TURN 88 89
FT HELIX 96 108
FT TURN 109 110
FT STRAND 113 115
FT TURN 120 123
FT TURN 130 131
FT HELIX 134 139
FT TURN 140 142
FT HELIX 145 147
FT HELIX 152 156
FT TURN 167 167
FT STRAND 170 175
FT TURN 178 179
FT HELIX 182 197
FT HELIX 218 223
FT TURN 224 225
FT HELIX 226 234
FT STRAND 235 236
FT TURN 242 243
FT STRAND 244 245
FT STRAND 248 255
FT TURN 257 258
FT STRAND 261 269
FT HELIX 273 276
FT TURN 277 277
FT TURN 281 282
FT STRAND 286 286
FT HELIX 287 295
FT HELIX 298 302
FT TURN 302 305
FT TURN 309 311
FT TURN 313 314
FT HELIX 316 317
FT HELIX 327 341
FT TURN 342 342
FT HELIX 345 363
FT STRAND 367 367
FT STRAND 376 376
FT HELIX 381 390
FT TURN 391 391
FT HELIX 394 402
FT HELIX 419 449
FT STRAND 458 464
FT STRAND 474 474
FT HELIX 476 506
FT TURN 507 508
FT HELIX 521 528
FT HELIX 536 543
FT HELIX 551 561
FT TURN 562 564
FT STRAND 566 570
FT STRAND 570 570
FT STRAND 579 584
FT TURN 585 585
FT STRAND 587 592
FT TURN 594 595
FT HELIX 596 601
FT HELIX 606 613
FT TURN 614 614
FT HELIX 618 623
FT TURN 624 624
FT HELIX 658 674
FT TURN 675 675
FT STRAND 677 684
FT TURN 692 693
FT HELIX 697 706
FT TURN 707 708
FT HELIX 709 718
FT STRAND 722 725
FT HELIX 726 733
```

```
FT HELIX 734 737
FT TURN 738 739
FT HELIX 748 759
FT TURN 760 760
FT HELIX 763 765
FT STRAND 766 768
FT STRAND 772 775
FT TURN 777 778
FT HELIX 779 787
SQ SEQUENCE 1978 AA, 228663 MW, B7B6C923E5273D93 CRC64;

Query Match 5.4%; Score 94; DB 1; Length 1978;
Best Local Similarity 23.5%; Pred. No. 50;
Matches 42; Conservative 41; Mismatches 72; Indels 24; Gaps 8;

QY 81 NISQFDEVD-----TDQGVLSPRE-IRT-LATRIHELPLSLDPLGLEMLIN 128
Db 1247 SLSQAKQDVHKKKKLEVLQDLSKYSQSGERVELNKKYKLGIEVENVSLNKEAS 1306
QY 129 CSMPLPADITQL-NIIPFOESYDPNLPVTKSLVTNCKPVTDKIHAKYKDKKXRPFI 187
Db 1307 KNIKLTQDVATLGSQLODTQEL-----LOEETROKLN---VTTKLRQLEDDKSLQ-EQ 1356
QY 188 MGEETIAFMRIRNVSHVVGOLDIRKNRKFCVCLNDNIDHNKDAQ-TVKAVLDFVE 245
Db 1357 LDEEVAZAKONLEHISTITQLSDSKKLQEFATVETMEBKKLQREIESLTOQFEB 1415

RESULT 15
SYN_BORBU STANDARD; PRT; 353 AA.
ID OS1038;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase)
DE (TrpRS).
GN TRPS OR TRSA OR BS0005.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.W., Casjens S., Huang W.M., Suton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gunn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D., Hanson M.,
RA Peterson J., Kierlavage A.R., Quackenbush J., Salzberg S.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman C.,
RA Uterback T., Matthey L., McDonald L., Artlich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RT Nature 390:580-586 (1997).
RL
RL Nature 390:580-586 (1997).
CC -I- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
CC diphosphate + L-tryptophanyl-tRNA(Trp).
CC -I- SUBUNIT: Homodimer (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001115; AAC66398.1; -.
CC PIR; E70100; E70100.
CC HSP; P00953; 1D2R.
```


This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:05:03 ; Search time 21.4324 Seconds
(without alignments)
4828.666 Million cell updates/sec

Title: US-10-657-280-2
Perfect score: 1731
Sequence: 1 DTFADSLRYVVKILNSKFGF.....RKIFPRRRRIHKEASPNRIRV 328

Scoring table: BLOSUM62

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

1:  sp archaea:*
2:  sp archaea:*
3:  sp archaea:*
4:  sp archaea:*
5:  sp archaea:*
6:  sp archaea:*
7:  sp archaea:*
8:  sp archaea:*
9:  sp archaea:*
10: sp archaea:*
11: sp archaea:*
12: sp archaea:*
13: sp archaea:*
14: sp archaea:*
15: sp archaea:*
16: sp archaea:*
17: sp archaea:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1731	100.0	450	4	Q86TQ2	Q86TQ2 homo sapien
2	1731	100.0	950	4	Q9UJL2	Q9UJL2 homo sapien
3	1290	74.5	298	4	Q9NPW9	Q9NPW9 homo sapien
4	704	40.7	132	4	Q9NT34	Q9NT34 homo sapien
5	581	39.6	324	1	Q8J340	Q8J340 mus musculus
6	506	29.2	66	1	Q8SK14	Q8SK14 drosophila
7	506	29.2	66	5	Q9V553	Q9V553 drosophila
8	113.5	6.6	55	2	Q68J15	Q68J15 neisseria m
9	113.5	6.6	55	2	Q84J00	Q84J00 neisseria m
10	113.5	6.6	55	2	Q84C29	Q84C29 neisseria m
11	113.5	6.6	55	2	Q83J59	Q83J59 neisseria m
12	113.5	6.6	54	16	Q9JW88	Q9JW88 neisseria m
13	111	6.4	1048	3	Q26O23	Q26O23 plasmodium
14	111	6.4	3119	5	Q25857	Q25857 plasmodium
15	111	6.2	3119	5	Q84T43	Q84T43 plasmodium
16	106.5	6.2	483	5	Q86K00	Q86K00 dictyosteli

17	106	6.1	574	16	Q9AK9	Q9AK9 streptococc
18	106	6.1	739	8	Q9MJ79	Q9MJ79 physarum po
19	106	6.1	2148	5	Q9BJD3	Q9BJD3 physarum po
20	105	6.0	666	13	Q9OXS4	Q9OXS4 xenopus lae
21	104.5	6.0	1999	11	Q63731	Q63731 rattus norri
22	104	6.0	524	5	Q8II88	Q8II88 plasmodium
23	104	6.0	574	16	Q8K839	Q8K839 streptococc
24	103	6.0	621	16	Q8I3R0	Q8I3R0 bacillus ce
25	103	6.0	620	17	Q57815	Q57815 pyrococcus
26	102.5	5.9	676	5	Q7YX23	Q7YX23 cytoprosperi
27	102.5	5.9	706	8	Q98S94	Q98S94 guillardia
28	102	5.9	515	3	Q99229	Q99229 saccharomyc
29	101.5	5.9	301	16	Q8Y9G9	Q8Y9G9 listeria m
30	101.5	5.9	408	12	Q9J575	Q9J575 fowlpox viru
31	101	5.8	375	10	Q7XMH9	Q7XMH9 oryza sativ
32	101	5.8	574	16	Q8P1M3	Q8P1M3 streptococc
33	100.5	5.8	265	16	Q8D241	Q8D241 wiglesworth
34	100.5	5.8	468	5	Q86BM6	Q86BM6 dirosophila
35	100	5.8	5987	5	Q8IDN3	Q8IDN3 plasmodium
36	99.5	5.7	893	4	Q9H0J3	Q9H0J3 homo sapien
37	99	5.7	723	16	Q8RM48	Q8RM48 fusobacteri
38	98	5.7	5687	5	Q8IHH4	Q8IHH4 plasmodium
39	97.5	5.6	893	16	Q51543	Q51543 borrelia bu
40	97.5	5.6	735	4	Q8IYX0	Q8IYX0 homo sapien
41	97.5	5.6	1498	16	Q9ZFP9	Q9ZFP9 listeria in
42	97	5.6	568	5	Q18679	Q18679 caenorhabdi
43	97	5.6	620	16	Q99Z27	Q99Z27 streptococc
44	97	5.6	746	16	Q8E8I6	Q8E8I6 streptomyc
45	97	5.6	942	16	Q9I1I4	Q9I1I4 streptomyc

ALIGNMENTS

Query Match	Best Local Similarity	Matches	Conservative	Score	DB	Length	Indels	Gaps
1	100.0%	328	0	1731	4	490	0	0
163	100.0%	0	0	128	0	0	0	0
61	100.0%	0	0	128	0	0	0	0
223	100.0%	0	0	128	0	0	0	0
121	100.0%	0	0	128	0	0	0	0

```

Db      283 GLEHMLINCSKMLPADITQLNNI PPTQESYVDNIPVTKSLVTNCKEPTDKIHKAYKDK 342
QY      181 NKYRPEIMGEELIAFMIRTNVSHVGGQDDIRKPKRFVCLNDIDHNHKAQIVKAVL 240
Db      343 NKYRPEIMGEELIAFMIRTNVSHVGGQDDIRKPKRFVCLNDIDHNHKAQIVKAVL 402
QY      241 RDPYEMFPIPSQFELPREYRNFLLMHLEQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 300
Db      403 RDPYEMFPIPSQFELPREYRNFLLMHLEQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 462
QY      301 AEQIALKRRKIPRRRIHKEASPNRIRV 328
Db      463 AEQIALKRRKIPRRRIHKEASPNRIRV 490

```

RESULT 2

```

Q9ULI2 PRELIMINARY; PRT; 950 AA.
AC Q9ULI2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA1208 (Fragment).
GN KIAA1208.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirosewa M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345 (1999).
DR EMBL; AB033034; BAA86522.2; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00036; ehand; 1.
DR Pfam; PF00066; notch; 1.
DR SMART; SMO0004; NT; 2.
DR PROSITE; PS00018; EF_HAND; 1.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 950 AA; 108785 MM; F653BDFAEAC7503 CRC64;

```

```

Query Match      100.0%; Score 1731; DB 4; Length 950;
Best Local Similarity 100.0%; Pred. No. 2.3e-128;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 DTFADSLRYVKNILNSKFGFTSRKYPAHMPHMDIRIVMQLQDMPPEEFDKTSFKYRHS 60
Db      623 DTFADSLRYVKNILNSKFGFTSRKYPAHMPHMDIRIVMQLQDMPPEEFDKTSFKYRHS 682
QY      61 EDMQPAFSYFYLYLMSAVQPLNISQVDEVDTPDQSGVLSDRERLTATRIHELPLSLDIT 120
Db      683 EDMQPAFSYFYLYLMSAVQPLNISQVDEVDTPDQSGVLSDRERLTATRIHELPLSLDIT 742
QY      121 GLEHMLINCSKMLPADITQLNNI PPTQESYVDNIPVTKSLVTNCKEPTDKIHKAYKDK 180
Db      743 GLEHMLINCSKMLPADITQLNNI PPTQESYVDNIPVTKSLVTNCKEPTDKIHKAYKDK 802
QY      181 NKYRPEIMGEELIAFMIRTNVSHVGGQDDIRKPKRFVCLNDIDHNHKAQIVKAVL 240
Db      803 NKYRPEIMGEELIAFMIRTNVSHVGGQDDIRKPKRFVCLNDIDHNHKAQIVKAVL 862
QY      241 RDPYEMFPIPSQFELPREYRNFLLMHLEQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 300

```

```

Db      863 RDPYEMFPIPSQFELPREYRNFLLMHLEQEMRAYRDKLKFWTHCVLATLIMFTIFSFF 922
QY      301 AEQIALKRRKIPRRRIHKEASPNRIRV 328
Db      923 AEQIALKRRKIPRRRIHKEASPNRIRV 950

```

RESULT 3

```

Q9NPM9 PRELIMINARY; PRT; 248 AA.
AC Q9NPM9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP762B26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RA Bloembergen H., Boecher M., Brandt P., Mewes H.W., Weill B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DDSI databases.
DR EMBL; AL559588; CAB94874.1; -.
DR PIR; T50618; T50618.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; 1.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 248 AA; 29336 MM; D4C4CA6C4505522 CRC64;

```

```

Query Match      74.5%; Score 1290; DB 4; Length 248;
Best Local Similarity 100.0%; Pred. No. 3.4e-94;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      85 VFPEVDTDSQVLSDEIRTLATRIHELPLSLQDITGLEHMLINCSKMLPADITQLNIP 144
Db      5 VFPEVDTDSQVLSDEIRTLATRIHELPLSLQDITGLEHMLINCSKMLPADITQLNIP 64
QY      145 PTOESYVDNIPVTKSLVTNCKEPTDKIHKAYKDKKRYRPFIMGEELIAFMIRTNVSH 204
Db      65 PTOESYVDNIPVTKSLVTNCKEPTDKIHKAYKDKKRYRPFIMGEELIAFMIRTNVSH 124
QY      205 VVGQDDIRKPKRFVCLNDIDHNHKAQIVKAVLRDPYEMFPIPSQFELPREYRNF 264
Db      125 VVGQDDIRKPKRFVCLNDIDHNHKAQIVKAVLRDPYEMFPIPSQFELPREYRNF 184
QY      265 LHMHELOEMRAYRDKLKFWTHCVLATLIMFTIFSFAEQIALKRRKIPRRRIHKEASPN 324
Db      185 LHMHELOEMRAYRDKLKFWTHCVLATLIMFTIFSFAEQIALKRRKIPRRRIHKEASPN 244
QY      325 RIRV 328
Db      245 RIRV 248

```

RESULT 4

```

Q9NV34 PRELIMINARY; PRT; 132 AA.
AC Q9NV34;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ10959.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```


RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuko Y., Oshima A.,
RT "NEO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001821; EAA91926.1; -;
SQ SEQUENCE 132 AA; 16083 MW; 92B5DB4E062AD07 CRC64;
KW Hypothetical protein.
Query Match 40.7%; Score 704; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 3,66-48;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 197 MIRNVSHVVGQDDIRKPKFVCINDHNHKAQTVXALRDPFESMFPIPSQPEL 256
DB 1 MIRNVSHVVGQDDIRKPKFVCINDHNHKAQTVXALRDPFESMFPIPSQPEL 60
QY 257 PREVENRPLHMEHLEOMRAYRDKLKEWTHCVATLIMFTISFFAEQILAKRKIFPPRR 316
DB 61 PREIKRFLHMELEOMRAYRDKLKEWTHCVATLIMFTISFFAEQILAKRKIFPPRR 120
QY 317 IHKEASPNRIRY 328
DB 121 IHKEASPNRIRY 132
RESULT 5
Q61340 PRELIMINARY; PRT; 384 AA.
AC Q61340;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Basic domain/leucine zipper transcription factor (Fragment).
GN MAFB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95094266; PubMed=8001130;
RA Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene *kr* encodes a novel basic domain-leucine
RT zipper transcription factor";
RL Cell 79:1025-1034(1994).
DR EMBL; U36434; AAA65688.1; -;
DR PIR; I49528; I49528.
DR MGD; MGI:104555; MafB.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:000357; P:regulation of transcription from Pol II pro. . . , IDA.
DR GO; GO:000379; P:segment specification; IMP.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 1.
DR PROSITE; PS00018; EF_HAND; 1.
FT NON_TER
SQ SEQUENCE 384 AA; 43754 MW; 51F473C8807A7E55 CRC64;
Query Match 33.6%; Score 581; DB 11; Length 384;
Best Local Similarity 97.4%; Pred. No. 7,3e-38;
Matches 112; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAMPHMIDRIVQELQDMPEEFDKTSFHKVRS 60
DB 267 DTFADSLRYVVKILNSKFGFTSRKVPAMPHMIDRIVQELQDMPEEFDKTSFHKVRS 326
QY 61 EDMQAFASFYFLMSAVQPLNISQVDEVDYTDQSGVLSDRERLTATRIHEPLSL 115
DB 327 EDMQAFASFYFLMSAVQPLNISQVDEVDYTDQSGVLSDRERLTATRIHEPLSL 381

RESULT 6
Q68X14 PRELIMINARY; PRT; 666 AA.
ID Q68X14
AC Q68X14;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE R85033P.
GN CG8027.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STEFAL=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuno U., Paoleo U., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceuliker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RU EMBL; AY089618; AL90356.1; -;
DR FLYBase; FBgn0033392; CG8027.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00066; notch; 1.
SQ SEQUENCE 666 AA; 77745 MW; F6FDB6D1C1C39248 CRC64;
Query Match 29.2%; Score 506; DB 5; Length 666;
Best Local Similarity 33.6%; Pred. No. 1,2e-31;
Matches 108; Conservative 70; Mismatches 133; Indels 10; Gaps 5;
QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAMPHMIDRIVQELQDMPEEFDKTSFHKVRS 60
DB 348 DIYSHSLATNMMLNAYGKARVLAHVGFLDKDIVEMQRRFQQLDTHGQFPAP 407
QY 61 EDMQAFASFYFLMSAVQPLNISQVDEVDYTDQSGVLSDRERLTATRIHEPLSL 120
DB 408 TDLQYAFAYVYSFILMSETKWMSVEIPEDEPTDGSATWSDEVRFTFLRIYQPLDWSAMR 467
QY 121 GLSHMLINCSK---MLPADITQLNNIPFOESYDPLPPVYKSLVTNCKPVTDKIHA 176
DB 468 YFEVYVQNCNRLNGMLKVDYVEHSTL--VYERYEBSNLPITRIIDLVAQPLLAELAAAN 525
QY 177 YKDRNKYREFIMGE--BELAFKIRNVSHVVGQDDIRKPKFVCLNDNIDHNF-KDA 233
DB 526 FAVRPKYNFVSPKRSKSHSNFMALTNLEVVESLRLRRNPKFKFCINDNDANGEEN 585
QY 234 QYKAVLRADYEEMFPIPSQPELPRRYRRLFLHMELEOMRAYRDKLKEWTHCVATLIM 293
DB 586 EMTVRLHLEDYLSLFFRRRSKFEELPQYRNRFBESMRDFOVKR-RKAAVLVIGVSLLV 644
QY 294 FTFSFPAEQILAKRKIFPR 314
DB 645 VCLLRFCVCHKAKLVRCVQR 665
RESULT 7
Q9V553 PRELIMINARY; PRT; 666 AA.
ID Q9V553
AC Q9V553;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG8027-PA.
GN CG8027.
OS Drosophila melanogaster (fruit fly).

Query Match	29.2%	Score 506	DB 5	Length 666
Best Local Similarity	33.6%	Pred. No. 1.2e-31		
Matches 108	Conservative 70	Mismatches 133	Indels 10	Gaps 5
QY	1	DTFADSLRVNKLINSKFGFTSRKRVPAHMPHMDIRIVMOELQDMFPEEPEDKTSNKKVRS	60	
DB	348	DIYSHSLIATNMTLNRAYGFKARVLVAHVGFLIDKIDIVEAMORAFHOQIIDTAHOREPAR	407	
QY	61	EDMOFAFSYFYLMSAAQPINISQVPEVPTDQSGVLSDEIRTLARHIELPLSLQDLT	120	
DB	408	TDLOYAAYYSFLMSERKMSVSEIIPFEFTDQSAIWSDEVRFTLRIQPLPDSAMR	467	
QY	121	GLEHMLINCSK---MLPADITQLNIPPOESYYDNPDPVTSLVNTCKPVTDKIHKA	176	
DB	468	YFEEVYQNCNTRNLGMHLKVDIVESHSL--YERYEDSNLEPTITRDLVVRCPLLAEALAN	525	
QY	177	YKDNKKRPFELMBG--BEIAFKMRTNVSYYVGQOLDIRKNPFRFVCLNINIDNH-KDA	233	
DB	526	FAVAPKPNFVSPKRTSHSPMLTSLUTEVESLDRKRNPKFKNCINDNDANREDN	585	
QY	234	QTVKAVLDVDEYSEMFPIPSQFELPREYRNFHLNHELQEMRAYRDKLKFTHCYATLIM	293	
DB	586	EMVHLLLEDYLSFPFPRSKFELPPQYRNFFESWRDPQWRK-RKRAVLVIGYGVSLIV	644	
QY	294	FTIFSFEQILAKRKIFPR	314	
DB	645	VCLRFMCHHKAKLVRCVQR	665	
RESULT 8				
ID	068215	PRELIMINARY;	PRT;	545 AA.
AC	068215			
DT	01-AUG-1998	(TrEMBLrel. 07, Created)		
DT	01-AUG-1998	(TrEMBLrel. 07, Last sequence update)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)		
DS	SACB.			
GN	Neisseria meningitidis (serogroup A).			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;			
OC	Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=65699;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PB229;			
RC	MEDLINE=98175678; PubMed=9515923;			
RX	Swartley J.S., Liu L.-J., Miller Y.K., Martin L.E., Edupuganti S.,			
RA	Stephens D.S.;			
RT	"Characterization of the gene cassette required for biosynthesis of			
RT	the (alpha1-5)-linked N-acetyl-D-mannosamine-1-phosphate capsule of			
RT	serogroup A Neisseria meningitidis.";			
RL	J. Bacteriol. 180:1533-1539(1998).			
DR	EMBL: AF019760; AAC38286.1; -			
SO	SEQUENCE 545 AA; 64097 MW; 1BC3400C5CC33DCD CRC64;			
Query Match	6.6%	Score 113.5;	DB 2;	Length 545;
Best Local Similarity	18.4%	Pred. No. 1.1;		
Matches 47	Conservative 36;	Mismatches 65;	Indels 107;	Gaps 8;
QY	3	PADSLRVNKLINSKFGFTSRKRVPAHMPHMDIRIVMOELQDMFPEEPEDKTSNKKVRS	62	
DB	395	YLNARVANTLLEKPKFTTKLHTSHPSQMRDILFEMKKYPEENRRLHNKKRSLDD	454	
QY	63	MQF-AFSYFYLLMSAVOPLNISQVPEVPTDQSGVLSDEIRTLARHIELPLSLQDLTG	121	
DB	455	IAVTGYLYHHY-----ALLSGRLQS-----SDKTE	480	
QY	122	LEHMLINCSKMLPADITQLNINIP-POESYYDNPDPVTSLVNTCKPVTDKIHKA	180	
DB	481	LVQGNHDFK-----KLNNAVTLTKERFQ-KLP-----	508	
QY	181	NKRFELMGEELIAFKMRTNVSYYVGQOLDIRKNPFRFVCLNINIDNHKDAQTVKAVL	240	

```

Db      509 -----LSVCINDGADSHLNEENVQVI- 530
Qy      241 RDPYBSMPPIPSQFE 255
          ||:|:|:|:|
          531 -KFLETLFPLPSSSFE 544
Db

RESULT 9
Q84D00
ID      Q84D00      PRELIMINARY;      PRT;      545 AA.
AC      Q84D00;
DT      01-JUN-2003 (TREMBlrel. 24, Created)
DT      01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      SAcB.
GN      SAcB.
OS      Neisseria meningitidis.
OC      Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC      Neisseriaceae; Neisseria.
OX      NCBI_TaxID=487;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=M4775;
RA      Sacchi C.T., Whitney A.A., Mayer L.W., Motherhead E., Popovic T.;
RT      "Neisseria meningitidis sacB gene.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY234202; AA085300.1; -.
SQ      SEQUENCE 545 AA; 64134 MW; A3648A3C85AA946F CRC64;

Query Match
Best Local Similarity 18.4%; Pred. No. 1.1;
Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

Qy      3 FADSLRYVVKILNSKGFSTRKVPAPMPHMDRIYVQELQDMPEEFDKTSFHKVHSHD 62
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          395 YINGARANTLLEKEFKFTTKLHTSPQSMRTDILFEMKKYPEEFNRTLHNRFRSLDD 454
Db
Qy      63 MGF-AFSFYIYMSAVQPINISQVFDEVDTDSGVLSDREIRTLATRIHELPLSLDPLTG 121
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          455 IAVTGYLYHHY-----ALLSGRALQS-----SDKTE 480
Db
Qy      122 LEHMLINCSKMLPADITQLNINPP-TOESYDPNLPPTKSLVTNCKPVTDKIKHAYKDK 180
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          481 LVQONHDFPK-----KLNNAVTLTKERNFD-KLP----- 508
Db
Qy      181 NKYRFEIMGEEBIAFKMIRTNVSHVQGLDIRKNRKFVCINDNDIDHNKDAQTVKAVL 240
          ||:|:|:|:|
          509 -----LSVCINDGADSHLNEENVQVI- 530
Db
Qy      241 RDPYBSMPPIPSQFE 255
          ||:|:|:|:|
          531 -KFLETLFPLPSSSFE 544
Db

RESULT 10
Q84C29
ID      Q84C29      PRELIMINARY;      PRT;      545 AA.
AC      Q84C29;
DT      01-JUN-2003 (TREMBlrel. 24, Created)
DT      01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      SAcB.
GN      SAcB.
OS      Neisseria meningitidis.
OC      Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC      Neisseriaceae; Neisseria.
OX      NCBI_TaxID=487;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=M2677;
RA      Sacchi C.T., Whitney A.A., Mayer L.W., Motherhead E., Popovic T.;
RT      "Neisseria meningitidis sacB gene.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

```

```

DR      EMBL; AY234204; AA085302.1; -.
SQ      SEQUENCE 545 AA; 64128 MW; 016B534F50D5DD71 CRC64;

Query Match
Best Local Similarity 18.4%; Pred. No. 1.1;
Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

Qy      3 FADSLRYVVKILNSKGFSTRKVPAPMPHMDRIYVQELQDMPEEFDKTSFHKVHSHD 62
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          395 YINGARANTLLEKEFKFTTKLHTSPQSMRTDILFEMKKYPEEFNRTLHNRFRSLDD 454
Db
Qy      63 MGF-AFSFYIYMSAVQPINISQVFDEVDTDSGVLSDREIRTLATRIHELPLSLDPLTG 121
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          455 IAVTGYLYHHY-----ALLSGRALQS-----SDKTE 480
Db
Qy      122 LEHMLINCSKMLPADITQLNINPP-TOESYDPNLPPTKSLVTNCKPVTDKIKHAYKDK 180
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          481 LVQONHDFPK-----KLNNAVTLTKERNFD-KLP----- 508
Db
Qy      181 NKYRFEIMGEEBIAFKMIRTNVSHVQGLDIRKNRKFVCINDNDIDHNKDAQTVKAVL 240
          ||:|:|:|:|
          509 -----LSVCINDGADSHLNEENVQVI- 530
Db
Qy      241 RDPYBSMPPIPSQFE 255
          ||:|:|:|:|
          531 -KFLETLFPLPSSSFE 544
Db

RESULT 11
Q83U59
ID      Q83U59      PRELIMINARY;      PRT;      545 AA.
AC      Q83U59;
DT      01-JUN-2003 (TREMBlrel. 24, Created)
DT      01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      SAcB.
GN      SAcB.
OS      Neisseria meningitidis.
OC      Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC      Neisseriaceae; Neisseria.
OX      NCBI_TaxID=487;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=M1124, and M3562;
RA      Sacchi C.T., Whitney A.A., Mayer L.W., Motherhead E., Popovic T.;
RT      "Neisseria meningitidis sacB gene.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=M7060;
RA      Sacchi C.T., Whitney A.W., Mayer L.W., Popovic T.;
RT      "Neisseria meningitidis sacB gene.";
RL      Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY234203; AA085301.1; -.
DR      EMBL; AY234205; AA085303.1; -.
DR      EMBL; AY281049; AAP34772.1; -.
SQ      SEQUENCE 545 AA; 64133 MW; 7D2564286C95CD3D CRC64;

Query Match
Best Local Similarity 18.4%; Pred. No. 1.1;
Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

Qy      3 FADSLRYVVKILNSKGFSTRKVPAPMPHMDRIYVQELQDMPEEFDKTSFHKVHSHD 62
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          395 YINGARANTLLEKEFKFTTKLHTSPQSMRTDILFEMKKYPEEFNRTLHNRFRSLDD 454
Db
Qy      63 MGF-AFSFYIYMSAVQPINISQVFDEVDTDSGVLSDREIRTLATRIHELPLSLDPLTG 121
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          455 IAVTGYLYHHY-----ALLSGRALQS-----SDKTE 480
Db
Qy      122 LEHMLINCSKMLPADITQLNINPP-TOESYDPNLPPTKSLVTNCKPVTDKIKHAYKDK 180
          : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
          481 LVQONHDFPK-----KLNNAVTLTKERNFD-KLP----- 508

```

```

QY 181 NKYREIMGEELAFMIRTNVSHVVGOLDIRKPKRVCUNDINDHNRKDAQTVKAVL 240
DB 509 -----LSVCINDGADSHLNEEMNVQVI- 530

QY 241 RDPYEMPEIPSOQE 255
DB 531 -KFLETLFPLPSSFE 544

RESULT 12
QJQWMB
ID QJQWMB PRELIMINARY; PRT; 545 AA.
AC QJQWMB;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative capsule biosynthesis protein.
GN SABC OR NMA0200.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=62491 / Serogroup A / Serotype 4A;
RX MEDLINE=2022256; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Tagatz K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajadurai M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506 (2000).
DR EMBL; AL62752; CAB83514.1; -.
DR PIR; C82014; C82014.
KW Complete proteome.
SQ SEQUENCE 545 AA; 64133 MW; 7D2564286C95CD3D CRC64;

Query Match
Best Local Similarity 18.4%; Pred. No. 1.1; Length 545;
Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

QY 3 FADSLRYVKNILNSKFGFTSRKVPAMPMIDRIYMOELQMFPEEPKTSFHKVRSSE 62
DB 395 YLNGARNANTLLEKEFKFTTKLHTSPQSWRTDILFMEKKYCPPEFNRIHLNKRSLDD 454
QY 63 MGF-APSYFYLMASVQPLNISQVFDVDTQSGVLSPREIRTLATRIHELPLSLQDTG 121
DB 455 IAVTGLYHYH-----ALLSGRALQS-----SDKTE 480
QY 122 LEHMLINCSKMLPADITQINNIP-TQESYDYPNLPVTKSLVTNCKEPTDKIRKAYDK 180
DB 481 LVQONHDFPK-----KLNNVLTTKERNFD-KLP----- 508
QY 181 NKYREIMGEELAFMIRTNVSHVVGOLDIRKPKRVCUNDINDHNRKDAQTVKAVL 240
DB 509 -----LSVCINDGADSHLNEEMNVQVI- 530

QY 241 RDPYEMPEIPSOQE 255
DB 531 -KFLETLFPLPSSFE 544

RESULT 13
Q26023 PRELIMINARY; PRT; 1048 AA.
AC Q26023;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

```

```

DE Hypothetical protein (Fragment).
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Alano P., Elliott J.F.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; M69147; AAA74653.1; -.
KW Hypothetical protein.
FT NON TER 1
FT NON TER 1048
SQ SEQUENCE 1048 AA; 126518 MW; 7AF051480FA22424 CRC64;

Query Match
Best Local Similarity 22.7%; Pred. No. 3.9; Length 1048;
Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;

QY 41 LQMFPEEDPKTSFHKVRSSEDMQFAFSYFYLMASVQPLNISQVFEVDVTDQSGVLS 99
DB 592 LEDIYME---KKKYKEYSKMRRISSMLDYEVNK---QIKENHYKVDI-----ISBH 639
QY 100 --REIR---TLATIHLEPLSLQ-----DLTGLEHMLINCSKMLPADITQINNIP-- 145
DB 640 KFGELIQHNRDKLENTHELHYKEMVQIQIDILNYYHOLENHSLELQALQONKNIPEHL 699
QY 146 -----TQESYDYPNLPV-TKSLVTNCKEPTDKIRKAYDK-----KMKYRFEI--- 187
DB 700 NVLEKLEITKR--KKNRPDISTSHARDEQVSTLIRGAHNGDIIKGDNDNEVLLI 757
QY 188 -----MGEELAFKIRITNVSHVVGQ-----LDDIRKPKRVCUNDINDH 228
DB 758 EQIQSLTKRWGQNVGSGILEKLNLSQYQYLODKLNIVVDIKNLNF--KHYLEK 814
QY 229 NHRDAO-----TVKAVRDPYEMPEIPSOQELPREYRNFLEHME--QEWKAVRDKLK 281
DB 815 LMKESKINERKFTIKVDVLSNYSTL-----EYWKVFL-LHDQGEWSEFEKDELE 862

RESULT 14
Q25857 PRELIMINARY; PRT; 3119 AA.
ID Q25857;
AC Q25857;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Pfg377.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=96360472; PubMed=8719156;
RA Alano P., Read D., Bruce M., Aikawa M., Kaide T., Tegoshi T.,
RA Bharti S., Smith D.K., Luo C., Hanra S., Carter R., Elliott J.F.;
RT "COS cell expression cloning of Pfg377, a Plasmodium falciparum
RT gametocyte antigen associated with oocystophilic bodies."
RL KOL. Biochem. Parasitol. 74:143-156 (1995).
DR EMBL; L04161; AAC37257.1; -.
DR PIR; T18414; T18414.
DR InterPro; IPR001064; Crystal11n.
DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
SQ SEQUENCE 3119 AA; 377353 MW; 519B92D5BDBFCRC CRC64;

Query Match
Best Local Similarity 22.7%; Pred. No. 15; Length 3119;
Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;

QY 41 LQMFPEEDPKTSFHKVRSSEDMQFAFSYFYLMASVQPLNISQVFEVDVTDQSGVLS 99
DB 1261 LEDIYME---KKKYKEYSKMRRISSMLDYEVNK---QIKENHYKVDI-----ISBH 1308

```

QY 100 --REIR-----TLATRIHELPLSLQ-----DITGLEHMLINCSKMLPADITQOLNNIP-- 145
 Db 1309 KFOELRQHRDKIENTHELKEMVVOIQIDLTNYHOLENISELLOALQONKNIPRHL 1368
 QY 146 -----TQESYDPLNPV-TKSLVTNCKPVTDKIHAKYD-----KMKRPEI--- 187
 Db 1369 NVLEKKLEITTKR--KKKKPDISTSSHATDEQVSDTLIRGAHNGDIIKGEDNDEVLLI 1426
 QY 188 -----MGEEBIAFKMIRTNVSHVVGQ-----LDDIRKNPRKFCVCLNDNIDH 228
 Db 1427 EGIQSLKTMGDNQVGSILEKLNLSPOQYOLQDKNLVVEDIYKNLRNF---KHVYIEK 1483
 QY 229 NHKDAO-----TYKAVLRDPFYESGFPIPSQFELPREYRNFELHMEHLOEWRAVYRDKLK 281
 Db 1484 LHKESKINREKFTTKVDVLSNVYSTL-----EYWKFL-LHDFQWSEFEKDELE 1531

RESULT 15

0814T3 PRELIMINARY; PRT: 3119 AA.
 ID 0814T3
 AC 0814T3
 DT 01-MAR-2003 (TEMBLrel. 23, Created)
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE PFG377 protein.
 GN PFL2405C.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=2255705; PubMed=12368664;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain N., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shalim S.J., Suh B., Peterson J., Angiuoli S.,
 RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrrell B.,
 RA "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.";
 RL Nature 419:498-511 (2002).
 DR EMBL; AE014852; AAN36565.1; -;
 DR InterPro; IPR001064; Crystal1in.
 DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
 SQ SEQUENCE 3119 AA; 377380 MW; 3A6B67FICD5C1CF CRC64;

Query Match 6.4%; Score 111; DB 5; Length 3119;
 Best Local Similarity 22.7%; Pred. No. 15;
 Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;

QY 41 LODMFEPEBDKTSFHKHSEDMQAFSGFYTLMSAVQPLNISQVDFDEVDTQDSGLSD- 99
 Db 1261 LEDITWE--KKKYKEEYSKMRRIISSNLDYVNR--QKEHYHKVDT-----ISEH 1308
 QY 100 --REIR-----TLATRIHELPLSLQ-----DITGLEHMLINCSKMLPADITQOLNNIP-- 145
 Db 1309 KFOELRQHRDKIENTHELKEMVVOIQIDLTNYHOLENISELLOALQONKNIPRHL 1368
 QY 146 -----TQESYDPLNPV-TKSLVTNCKPVTDKIHAKYD-----KMKRPEI--- 187
 Db 1369 NVLEKKLEITTKR--KKKKPDISTSSHATDEQVSDTLIRGAHNGDIIKGEDNDEVLLI 1426
 QY 188 -----MGEEBIAFKMIRTNVSHVVGQ-----LDDIRKNPRKFCVCLNDNIDH 228
 Db 1427 EGIQSLKTMGDNQVGSILEKLNLSPOQYOLQDKNLVVEDIYKNLRNF---KHVYIEK 1483
 QY 229 NHKDAO-----TYKAVLRDPFYESGFPIPSQFELPREYRNFELHMEHLOEWRAVYRDKLK 281

Db 1484 LHKESKINREKFTTKVDVLSNVYSTL-----EYWKFL-LHDFQWSEFEKDELE 1531

Search completed: July 26, 2004, 11:11:37
 Job time : 23.4324 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:28 ; Search time 28.1358 Seconds
(without alignments)
3062.893 Million cell updates/sec

Title: US-10-657-280-3
Perfect score: 1634
Sequence: 1 MAAGTARLLLLGSGAGPA.....TPRAKSPEQLRGDPGLRGSL 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1634	100.0	305	4	ABR61378 Human GLC
2	1634	100.0	305	5	ABG92075 Human rec
3	1634	100.0	305	6	ADA57553 Human sec
4	1634	100.0	305	6	ADA57217 Human sec
5	1634	100.0	305	6	ADA41096 Human sec
6	1634	100.0	305	6	ADA41445 Human sec
7	1634	100.0	305	7	ADB91814 Human sec
8	1634	100.0	305	7	ADB91665 Human sec
9	1634	100.0	305	7	ADC74325 Human sec
10	1634	100.0	305	7	ADC74565 Human sec
11	1634	100.0	305	7	ADD27815 Human GLC
12	1634	100.0	305	7	ABW01490 Human GLC
13	1634	100.0	305	7	ADD37973 Human sec
14	1634	100.0	305	7	ADD38084 Human sec
15	1634	100.0	305	7	ABW01539 Human GLC
16	1634	100.0	305	7	ADD45631 Human pro
17	1634	100.0	305	2	ADA36229 Human sec
18	1634	100.0	306	6	ADA11624 Human nov
19	1634	100.0	310	4	AAK25804 Human pro
20	1288.5	78.9	307	4	ABR61381 Murine GL
21	1288.5	78.9	307	7	ADD27820 GLCNAc-ph
22	1288.5	78.9	307	7	ABW01493 Mouse pro
23	1288.5	78.9	307	7	ABW01542 Mouse pro
24	1244.5	76.2	296	2	AA48273 Human pro
25	1092	66.8	230	7	AA030426 Human sec

26	495.5	30.3	94	2	AAV73905	AAV73905 Human pro
27	171	10.5	647	3	AAQ37772	AAQ37772 Arabidops
28	171	10.5	647	3	AAQ30910	AAQ30910 Arabidops
29	168	10.3	350	3	AAQ27315	AAQ27315 Arabidops
30	168	10.3	358	3	AAQ27509	AAQ27509 Arabidops
31	155.5	9.5	597	7	ADB89035	ADB89035 Human CGD
32	155.5	9.5	612	6	ABG74699	ABG74699 Human CGD
33	155.5	9.5	638	6	ABG74700	ABG74700 Human CGD
34	155.5	9.5	652	6	ABG74695	ABG74695 Human CGD
35	148	9.1	548	4	ABR62383	ABR62383 Drosophila
36	140.5	8.6	126	6	ABU70630	ABU70630 Human adi
37	140.5	8.6	524	4	AAU28043	AAU28043 Novel hum
38	140.5	8.6	527	4	AAU28042	AAU28042 Novel hum
39	140.5	8.6	527	4	ABR50266	ABR50266 Protein k
40	140.5	8.6	534	4	AAU28044	AAU28044 Novel hum
41	137.5	8.4	109	5	AAQ66098	AAQ66098 Wheat alp
42	125	7.6	483	4	AAQ33456	AAQ33456 Human pro
43	124.5	7.6	559	7	ADB80850	ADB80850 Novel pro
44	120.5	7.4	483	3	AAQ08639	AAQ08639 A murine
45	120.5	7.4	483	3	AAQ08647	AAQ08647 A murine

ALIGNMENTS

RESULT 1	
ABR61378	ABR61378 standard; protein; 305 AA.
ID	ABR61378;
XX	
AC	ABR61378;
XX	
DT	01-AUG-2003 (first entry)
XX	
DE	Human GLCNAc-phosphotransferase gamma-subunit.
XX	
KM	Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
KW	GLCNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;
KM	N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;
KW	enzyme replacement therapy; phosphorylated lysosomal hydrolase;
KM	lysosomal storage disease; enzyme; gamma-subunit.
XX	
OS	Homo sapiens.
XX	
FX	
FT	Key
FT	Peptide
FT	Protein
FT	
XX	
PN	US6537785-B1.
XX	
PD	25-MAR-2003.
XX	
PF	10-AUG-2000; 2000US-00636077.
XX	
PR	14-SEP-1999; 99US-015831P.
XX	
PA	(GENZ-) GENZYME GLYCOTOLOGY RES INST INC.
XX	
PI	Canfield WM;
XX	
DR	WPI; 2001-290356/30.
XX	
PT	N-PSDB; ACC81002.
XX	
PT	Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase; useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.
PS	Disclosure; Page 29-30; 62pp; English.
XX	
CC	The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GLCNAc-phosphotransferase) (I) and phosphodiester

CC alpha-glucosidase (N-acetylglucosamine-1-phosphodiester alpha-N-
CC Acetylglucosaminidase) (II). The protein of the invention has
CC nephrotoxic activity, and may be useful in enzyme replacement therapy. A
CC protein of the invention (I), (II) is useful for preparing a
CC phosphoylated lysosomal hydrolase. The phosphoylated hydrolase
CC comprising a terminal mannose-6-phosphate, is useful for treating a
CC patient suffering from a lysosomal storage disease. The present sequence
CC is used in the exemplification of the invention

XX SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLRLILLIGLSAGPAPAGAAKMKVVEBNPFGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLRLILLIGLSAGPAPAGAAKMKVVEBNPFGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVLHFLSGKCSLVSTYKYFCPPHNTQHEQTFRNAYSGIIGIHWHEMIANNPTTG 120
DB 61 PVLHFLSGKCSLVSTYKYFCPPHNTQHEQTFRNAYSGIIGIHWHEMIANNPTTG 120
QY 121 MMRDGDACRSRSRQSVLACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVPTL 180
DB 121 MMRDGDACRSRSRQSVLACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVPTL 180
QY 181 PEALOROMQVEODLADDELITPQGHKRLRTLPEDAGYLKTBENEPRLQEGGDSLGFE 240
DB 181 PEALOROMQVEODLADDELITPQGHKRLRTLPEDAGYLKTBENEPRLQEGGDSLGFE 240
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTPETSNLEHLGHETPRAXSPEDLRGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTPETSNLEHLGHETPRAXSPEDLRGDPG 300
QY 301 LKGS 305
DB 301 LKGS 305

RESULT 2
ABG92075 standard; protein; 305 AA.

XX AC ABG92075;
XX DT 29-NOV-2002 (first entry)

DE Human receptors and membrane associated protein REMAP-18.

XX KW Human; receptor and membrane-associated protein; REMAP; atherosclerosis;
KW cardiovascular disorder; hypertension; congestive heart failure; edema;
KW aneurysm; angina pectoris; ischaemic heart disease; lung disease; nausea;
KW rheumatic heart disease; chronic obstructive pulmonary disease; anaemia;
KW emphysema; bronchitis; gastrointestinal disorder; peptic ulcer; asthma;
KW Crohn's disease; lipid metabolism disorder; Fabry's disease; dementia;
KW pancreatitis; diabetes mellitus; hyperlipidaemia; autoimmune disorder;
KW inflammatory disorder; neurological disorder; Creutzfeldt-Jakob disease;
KW cell proliferative disorder; protein replacement therapy; adenocarcinoma;
KW developmental disorder; metabolic disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; multiple sclerosis; Cushing's syndrome; anxiety;
KW schizophrenia; Addison's disease; endocrine disorder; gene therapy; gout;
KW amnesia; cancer; leukaemia; lymphoma; myeloma; sarcoma.

OS Homo sapiens.
XX WO200263006-A2.
XX 15-AUG-2002.
XX PD 05-FEB-2002; 2002WO-US003866.
XX PF 06-FEB-2001; 2001US-0267201P.
XX PR

PR 16-FEB-2001; 2001US-0269580P.
PR 03-APR-2001; 2001US-0282679P.
PR 02-MAY-2001; 2001US-0288295P.
PR 14-JAN-2002; 2002US-0348687P.

PA (INCY-) INCYTE GENOMICS INC.

XX PI Aizawai Y, Yue H, Ding L, Nguyen DB, Gandhi AR, Burford N;
PI Thangavelu K, Elliott VS, Ramkumar U, Yao MG, Lal PG, Tang TY,
PI Swarnakar A, Warren BA, Walla NK, Policky JL, Xu Y, Honchell CD;
PI Au-Young J, Baughn ME, Duggan BM, Lu DAM, Gietzen KU, Hillmann JU;
PI Raumann BE, Lu Y, Sappenstein SK, Tran UK, Richardson TW;
PI Emerling BM, Hafalia Aza, Burrill JD, Marcus GA, Zingler KA;
PI Kable AE, Gorrard AE;

XX WPI, 2002-627559/67.
XX N-PSDB; ABS67796.

PT New human receptor and membrane-associated proteins and polynucleotides,
PT useful for diagnosing, treating or preventing cardiovascular disorders
PT (e.g. aneurysms), neurological disorders (e.g. Parkinson's disease) or
PT cancers.

XX Claim 2, Page 185-186; 262pp; English.

XX The present invention relates to a new receptor and membrane-associated
XX protein (REMAP). The polypeptide, polynucleotide and agonist are useful
XX for treating a condition associated with decreased expression of
XX functional REMAP. The antagonist is useful for treating a disease
XX associated with overexpression of functional REMAP. The anti-REMAP
XX antibody is useful for diagnosing a condition or disease associated with
XX the expression of REMAP. These polypeptides, polynucleotides, agonists
XX and antagonists are particularly useful for diagnosing, treating or
XX preventing cardiovascular (e.g. atherosclerosis, hypertension, aneurysms,
XX congestive heart failure, angina pectoris, or ischemic or rheumatic
XX heart disease), lung (e.g. edema, chronic obstructive pulmonary disease,
XX emphysema or bronchitis) gastrointestinal (e.g. nausea, peptic ulcer or
XX Crohn's disease), lipid metabolism (e.g. Fabry's disease, diabetes
XX mellitus or hyperlipidaemia), autoimmune/inflammatory (e.g. anaemia,
XX asthma, gout, pancreatitis or Crohn's disease), neurological (e.g.
XX dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
XX multiple sclerosis, Parkinson's disease, anxiety, schizophrenia or
XX amnesia), metabolic (e.g. Addison's disease), developmental (e.g.
XX Cushing's syndrome), endocrine or cell proliferative disorders (e.g.
XX cancers including adenocarcinoma, leukaemia, lymphoma, myeloma or
XX sarcoma). The present amino acid sequence represents a human REMAP
XX protein of the invention

XX SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 5; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLRLILLIGLSAGPAPAGAAKMKVVEBNPFGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLRLILLIGLSAGPAPAGAAKMKVVEBNPFGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVLHFLSGKCSLVSTYKYFCPPHNTQHEQTFRNAYSGIIGIHWHEMIANNPTTG 120
DB 61 PVLHFLSGKCSLVSTYKYFCPPHNTQHEQTFRNAYSGIIGIHWHEMIANNPTTG 120
QY 121 MMRDGDACRSRSRQSVLACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVPTL 180
DB 121 MMRDGDACRSRSRQSVLACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVPTL 180
QY 181 PEALOROMQVEODLADDELITPQGHKRLRTLPEDAGYLKTBENEPRLQEGGDSLGFE 240
DB 181 PEALOROMQVEODLADDELITPQGHKRLRTLPEDAGYLKTBENEPRLQEGGDSLGFE 240
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTPETSNLEHLGHETPRAXSPEDLRGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTPETSNLEHLGHETPRAXSPEDLRGDPG 300

QY 301 LRGS 305
 DB 301 LRGS 305

RESULT 3
 ADA57553
 ID ADA57553 standard; protein; 305 AA.
 AC ADA57553;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human secreted protein #500.
 XX
 KW immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
 KW cytoprotective; cerebroprotective; neuroprotective; nootropic;
 KW cardiovascular; antiarteriosclerotic; gene therapy;
 KW human secreted protein; immune disorder; inflammation;
 KW inflammatory bowel disease; cancer; CNS disorder; neurodegenerative disorders;
 KW multiple sclerosis; ischaemic brain injury; Parkinson's disease; asthma; allergy;
 KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
 KW triple helix formation; antisense gene therapy; forensic biology.
 XX
 OS Homo sapiens.
 XX
 PN WO2002102994-A2.
 XX
 PD 27-DEC-2002.
 XX
 PF 19-MAR-2002; 2002WO-US008278.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI, 2003-167512/16.
 DR N-PSDB; ADA56660.
 XX
 PT New human secreted polypeptides and polynucleotides, useful for
 PT diagnosing, treating or preventing e.g. immune disorders, inflammatory
 PT conditions, respiratory disorders, cancers, CNS disorders, or
 PT neurodegenerative disorders.
 XX
 PS Claim 13; SEQ ID NO 1746; 1754pp; English.
 XX
 CC The invention relates to 592 new human secreted polypeptides useful for
 CC diagnosing, treating or preventing e.g. immune disorders, inflammatory
 CC conditions, respiratory disorders, cancers, CNS disorders, or
 CC neurodegenerative disorders, or polypeptides comprising an amino acid
 CC sequence at least 98% identical to the new sequences. The polypeptides,
 CC antibodies or antibody fragments that bind to the polypeptides, nucleic
 CC acids encoding the polypeptides, agonists or antagonists that binds to
 CC the polypeptide, are useful in preparing diagnostic or pharmaceutical
 CC compositions for diagnosing, treating or preventing an e.g. immune
 CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,
 CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
 CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
 CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
 CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and
 CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
 CC polynucleotides are useful for chromosome identification, chromosome
 CC mapping, for controlling gene expression through triple helix formation
 CC or antisense DNA or RNA, in gene therapy, for identifying individuals
 CC from minute biological samples, in forensic biology, and as hybridization
 CC probes. The polypeptides are useful for as molecular weight markers on
 CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

CC gels, to raise antibodies, for testing biological activities, and for
 CC treating or preventing neural disorders, immune system disorders,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, proliferative and/or cancerous diseases. This sequence corresponds
 CC to one of the polypeptide of the invention. Note: The sequence data for
 CC this patent did form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 305 AA;
 SQ
 Query Match 100.0%; Score 1634; DB 6; Length 305;
 Best Local Similarity 100.0%; Pred. No. 3.2e-161;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLGLSAGPAPGAAKMKVEEPNAGVNNPPLPQASRLQAKRDPSPVG 60
 DB 1 MAAGLARLLLLGLSAGPAPGAAKMKVEEPNAGVNNPPLPQASRLQAKRDPSPVG 60
 QY 61 PVHLFPLSGKCSLVSTYKFCPPHNTQHEQTPRNAYSGIGIWHWEIANTFTG 120
 DB 61 PVHLFPLSGKCSLVSTYKFCPPHNTQHEQTPRNAYSGIGIWHWEIANTFTG 120
 QY 121 MMWRDGDACRSRSGKVEIACGKSNRLAHVSEPTCYVALFETPLVCHPHALLVPTL 180
 DB 121 MMWRDGDACRSRSGKVEIACGKSNRLAHVSEPTCYVALFETPLVCHPHALLVPTL 180
 QY 181 PEALGROMQVBDLADDELITPOGHEKILRTLPEDGKYLTPENEPPLLEGPSLGE 240
 DB 181 PEALGROMQVBDLADDELITPOGHEKILRTLPEDGKYLTPENEPPLLEGPSLGE 240
 QY 241 TLENCRAKHELSKEIKRLKGLITQGIPTPTPTSENLHGHETPRAKSPBOLRGDPG 300
 DB 241 TLENCRAKHELSKEIKRLKGLITQGIPTPTPTSENLHGHETPRAKSPBOLRGDPG 300
 QY 301 LRGS 305
 DB 301 LRGS 305

RESULT 4
 ADA57217
 ID ADA57217 standard; protein; 305 AA.
 AC ADA57217;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human secreted protein #500.
 XX
 KW immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;
 KW cytoprotective; cerebroprotective; neuroprotective; nootropic;
 KW cardiovascular; antiarteriosclerotic; gene therapy;
 KW human secreted protein; immune disorder; inflammation;
 KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
 KW inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
 KW multiple sclerosis; ischaemic brain injury; Parkinson's disease;
 KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
 KW triple helix formation; antisense gene therapy; forensic biology.
 XX
 OS Homo sapiens.
 XX
 PN WO2002102994-A2.
 XX
 PD 27-DEC-2002.
 XX
 PF 19-MAR-2002; 2002WO-US008278.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;
 XX MPI: 2003-167512/16.
 DR N-PSDB; ADA56321.
 XX
 PT New human secreted polypeptides and polynucleotides, useful for
 PT diagnosing, treating or preventing e.g. immune disorders, inflammatory
 PT conditions, respiratory disorders, cancers, CNS disorders, or
 PT neurodegenerative disorders.
 XX
 PS Claim 13; SEQ ID NO 1407; 1754bp; English.
 XX
 CC The invention relates to 592 new human secreted polypeptides useful for
 CC diagnosing, treating or preventing e.g. immune disorders, inflammatory
 CC conditions, respiratory disorders, cancers, CNS disorders, or
 CC neurodegenerative disorders, or polypeptides comprising an amino acid
 CC sequence at least 95% identical to the new sequences. The polypeptides,
 CC antibodies or antibody fragments that bind to the polypeptides, nucleic
 CC acids encoding the polypeptides, agonists or antagonists that binds to
 CC the polypeptide, are useful in preparing diagnostic or pharmaceutical
 CC compositions for diagnosing, treating or preventing an e.g. immune
 CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,
 CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
 CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
 CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
 CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and
 CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
 CC polynucleotides are useful for chromosome identification, chromosome
 CC mapping, for controlling gene expression through triple helix formation
 CC or antisense DNA or RNA, in gene therapy, for identifying individuals
 CC from minute biological samples, in forensic biology, and as hybridization
 CC probes. The polypeptides are useful for as molecular weight markers on
 CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
 CC gels, to raise antibodies, for testing biological activities, and for
 CC treating or preventing neural disorders, immune system disorders,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, proliferative and/or cancerous diseases. This sequence corresponds
 CC to one of the polypeptide of the invention. Note: The sequence data for
 CC this patent did form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 305 AA;
 Query Match 100.0%; Score 1634; DB 6; Length 305;
 Best Local Similarity 100.0%; Pred. No. 3.2e-161;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
 ID ADA41096
 AC ADA41096;
 XX
 AC ADA41096;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human secreted protein.
 XX
 KW Human; secreted protein; cancer; hyperproliferative disorder;
 KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
 KW anaemia; allergic reaction; asthma; cardiovascular disorder;
 KW wound healing; cytotoxic; immunosuppressive; neotropic; neuroprotective;
 KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
 KW vulneary; cardiant; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO2002102993-A2.
 XX
 PD 27-DEC-2002.
 XX
 PF 19-MAR-2002; 2002WO-US008123.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX MPI: 2003-175238/17.
 DR
 XX
 PT New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating cancer or other hyperproliferative disorder,
 PT asthma, allergies or AIDS.
 XX
 XX
 PS Claim 1; SEQ ID NO 1478; 3205bp; English.
 XX
 CC The invention relates to novel genes ADA3629-ADA4056 and proteins
 CC ADA4056-ADA41501 for human secreted proteins, useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
 CC fragments, and agonists or antagonists that bind to the polypeptide are
 CC useful for preparing a diagnostic or pharmaceutical composition for
 CC diagnosing or treating cancer or other hyperproliferative disorder. The
 CC polypeptides and nucleic acid molecules are also useful for detecting,
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
 CC or other hyperproliferative disorders including neoplasms, autoimmune
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
 CC thrombocytopenia), allergic reactions including asthma or eczema,
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
 CC fungal or viral infections including HIV/AIDS), or wound healing and
 CC disorders of epithelial cell proliferation. The nucleic acids are also
 CC useful for chromosome identification, radiation hybrid mapping or long-
 CC range restriction mapping, as molecular weight markers, or as
 CC hybridization or diagnostic probes. The polypeptides and antibodies are
 CC useful for providing immunological probes for differential identification
 CC of the tissues immunohistochemistry assays. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 6; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLGLSAGGPAPAGAAKMKVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLGLSAGGPAPAGAAKMKVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNVTOHEQTFRNNAVSGILGIWHEWEIANNFTG 120
DB 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNVTOHEQTFRNNAVSGILGIWHEWEIANNFTG 120
QY 121 MMNRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVYPTL 180
DB 121 MMNRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVYPTL 180
QY 181 PEALQROMDQVEODLDELITPOGHEKILRTLFEAGYLTKEPENEPTOLEGGPDSIGFE 240
DB 181 PEALQROMDQVEODLDELITPOGHEKILRTLFEAGYLTKEPENEPTOLEGGPDSIGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTGIGIPYTRPETSNEHLGHETPRAKSPQOLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTGIGIPYTRPETSNEHLGHETPRAKSPQOLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 6
ADA41445
ID ADA41445 standard; protein; 305 AA.

AC ADA41445;
DT 20-NOV-2003 (first entry)
XX
DE Human secreted protein.
XX
KW Human; secreted protein; cancer; hyperproliferative disorder;
KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
KW anaemia; allergic reaction; asthma; cardiovascular disorder;
KW wound healing; cytostatic; immunosuppressive; nocotropic; neuroprotective;
KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
KW vulnerability; cardiac; gene therapy.
XX
OS Homo sapiens.
XX
FN WO2002102993-A2.
XX
PD 27-DEC-2002.
XX
PF 19-MAR-2002; 2002WO-US008123.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-175238/17.
XX
XX New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosis,
PT preventing or treating cancer or other hyperproliferative disorder,
PT asthma, allergies or AIDS.
XX
PS Claim 1; SEQ ID NO 1828; 3205bp; English.
XX
XX The invention relates to novel genes ADA93629-ADA40565 and proteins

CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 6; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLGLSAGGPAPAGAAKMKVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLGLSAGGPAPAGAAKMKVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNVTOHEQTFRNNAVSGILGIWHEWEIANNFTG 120
DB 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNVTOHEQTFRNNAVSGILGIWHEWEIANNFTG 120
QY 121 MMNRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVYPTL 180
DB 121 MMNRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPALLVYPTL 180
QY 181 PEALQROMDQVEODLDELITPOGHEKILRTLFEAGYLTKEPENEPTOLEGGPDSIGFE 240
DB 181 PEALQROMDQVEODLDELITPOGHEKILRTLFEAGYLTKEPENEPTOLEGGPDSIGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTGIGIPYTRPETSNEHLGHETPRAKSPQOLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTGIGIPYTRPETSNEHLGHETPRAKSPQOLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 7
ADB91814
ID ADB91814 standard; protein; 305 AA.

AC ADB91814;
DT 04-DEC-2003 (first entry)
XX
DE Human secreted protein #SEQ ID 760.
XX
KW Secreted protein; gene therapy; antidiabetic; diabetes; human.
XX
OS Homo sapiens.
XX

PN WO2003004622-A2.
XX
PD 16-JAN-2003.
XX
PF 19-MAR-2002; 2002WO-US008124.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-229407/22.
XX
PT Nucleic acid encoding a human secreted protein is useful in diagnosing or
XX treating diabetes or conditions related to diabetes.
XX
PS Claim 3; SEQ ID NO 760; 1537bp; English.
XX
CC The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 305 AA;
Query Match 100.0%; Score 1634; DB 7; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPPLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPPLPQASRLQAKRDPSPVSG 60
QY 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPPLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPPLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIHEWEIANNPTFG 120
DB 61 PVHLFRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIHEWEIANNPTFG 120
QY 121 MMRRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL 180
DB 121 MMRRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL 180
QY 181 PEALOROMQOVQODLADDELITPQGEKILRTLFEDAGYIKTPEENEPQLBGGPDSLGE 240
DB 181 PEALOROMQOVQODLADDELITPQGEKILRTLFEDAGYIKTPEENEPQLBGGPDSLGE 240
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPPAKSPQOLRGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPPAKSPQOLRGDPG 300
QY 301 LRGSL 305
DB 301 LRGSL 305
RESULT 8
ADB91665
ID ADB91665 strand; protein; 305 AA.
XX

AC ADB91665;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human secreted protein #SEQ ID 611.
XX
KM Secreted protein; gene therapy; antidiabetic; diabetes; human.
XX
CS Homo sapiens.
XX
PN WO2003004622-A2.
XX
PD 16-JAN-2003.
XX
PF 19-MAR-2002; 2002WO-US008124.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-229407/22.
XX
PT Nucleic acid encoding a human secreted protein is useful in diagnosing or
XX treating diabetes or conditions related to diabetes.
XX
PS Claim 3; SEQ ID NO 611; 1537bp; English.
XX
CC The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in
CC identifying a binding partner by contacting the polypeptide with a
CC binding partner, and determining whether the binding partner increases or
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
CC antibody or its fragment, agonist or antagonist are useful for preparing
CC a pharmaceutical composition for diagnosing or treating diabetes or
CC conditions related to diabetes. The present sequence is that of the human
CC immunoglobulin Fc portion used to generate fusion proteins, increasing
CC the stability of the fused protein as compared to the secreted protein
CC only. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 305 AA;
Query Match 100.0%; Score 1634; DB 7; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPPLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPPLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIHEWEIANNPTFG 120
DB 61 PVHLFRLSGKCSLVSTYKYEFCEPHNVTQHEQTFRNMAVSGIIGIHEWEIANNPTFG 120
QY 121 MMRRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL 180
DB 121 MMRRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL 180
QY 181 PEALOROMQOVQODLADDELITPQGEKILRTLFEDAGYIKTPEENEPQLBGGPDSLGE 240
DB 181 PEALOROMQOVQODLADDELITPQGEKILRTLFEDAGYIKTPEENEPQLBGGPDSLGE 240
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPPAKSPQOLRGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPPAKSPQOLRGDPG 300

QY 301 LRGS 305
Db 301 LRGS 305

RESULT 9
ADCT4325
ID ADC74325 strand; protein; 305 AA.
AC ADC74325;
XX 01-JAN-2004 (first entry)
XX Human secreted protein - SEQ ID 958.

KM anti-anemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;
KM antidiabetic; immunosuppressive; dermatological; nephrotropic;
KM antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;
KM fungicide; antiparasitic; antiarteriosclerotic; vulnery; cytostatic;
KM haemopoietic; haematologic; anaemia; autoimmune disorder;
KM rheumatoid arthritis; inflammation; Grave's disease; diabetes;
KM systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
KM Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
KM cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
human.

OS Homo sapiens.
XX MO2003038063-A2.
XX 08-MAY-2003.
XX 19-MAR-2002; 2002WO-US008277.
XX 21-MAR-2001; 2001US-0277340P.
XX 19-JUL-2001; 2001US-0306171P.
XX 13-NOV-2001; 2001US-0331287P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI: 2003-430516/40.
XX N-PSDB; ADC73710.
XX New human secreted polypeptide for diagnosing, preventing or treating
PT hematopoietic or hematologic disorders (e.g. anemia), autoimmune
PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or
PT atherosclerosis).

PS Claim 16; SEQ ID NO 958; 2272BP; English.
XX The invention relates to a novel human secreted polypeptide comprising a
CC defined sequence given in the specification. The polypeptide, nucleic
CC acid molecule, antibody, agonist or antagonist of the invention may be
CC useful for preparing a composition for diagnosing or treating a
CC haemopoietic or haematologic disorder such as anaemia, autoimmune
CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,
CC diabetes, systemic lupus erythematosus or glomerulonephritis,
CC neurodegenerative disorders including Parkinson's disease and Alzheimer's
CC disease, wounds and hyperproliferative disorders including
CC atherosclerosis or cancer, as well as bacterial, viral, fungal or
CC parasitic infections. The polypeptide may also be used during gene
CC therapy procedures and for identifying a binding partner by contacting
CC the polypeptide with a binding partner and determining whether the
CC binding partner increases or decreases the activity of the polypeptide.
CC The current sequence is that of the human secreted protein of the
CC invention.
XX Sequence 305 AA;
SQ

Query Match 100.0%; Score 1634; DB 7; Length 305;

Best Local Similarity 100.0%; Pred. No. 3.2e-161;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARIILLILGLSAGGAPAGAAKXKVEEENAFGVNPPFLQASRLQAKRDPSPVSG 60
Db 1 MAAGLARIILLILGLSAGGAPAGAAKXKVEEENAFGVNPPFLQASRLQAKRDPSPVSG 60

QY 61 PVHLFRLSGKCFSLVSTKYERCFPHNTQHEQTFRMNAVSGILGIMHEMELANNTPFG 120
Db 61 PVHLFRLSGKCFSLVSTKYERCFPHNTQHEQTFRMNAVSGILGIMHEMELANNTPFG 120

QY 121 MMNRDGDACSRSRQSKVELACGKSNRLAHVSEPTCYVALTFEPFLVCHPALVYPTL 180
Db 121 MMNRDGDACSRSRQSKVELACGKSNRLAHVSEPTCYVALTFEPFLVCHPALVYPTL 180

QY 181 PEALQROMQVPEDDLDELITPGHEKRLRTLFEDEGYLKTPEENPTQLGGPSLGE 240
Db 181 PEALQROMQVPEDDLDELITPGHEKRLRTLFEDEGYLKTPEENPTQLGGPSLGE 240

QY 241 TLNCRKAHKELSKEIKRLKGLLTQGIPIYTRPTESNLEHGHETPRAKSPQQLRGDPG 300
Db 241 TLNCRKAHKELSKEIKRLKGLLTQGIPIYTRPTESNLEHGHETPRAKSPQQLRGDPG 300

QY 301 LRGS 305
Db 301 LRGS 305

RESULT 10
ADCT4565
ID ADC74565 standard; protein; 305 AA.
XX 08-MAY-2003.
XX 01-JAN-2004 (first entry)
XX Human secreted protein - SEQ ID 1198.

XX anti-anemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;
KM antidiabetic; immunosuppressive; dermatological; nephrotropic;
KM antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;
KM fungicide; antiparasitic; antiarteriosclerotic; vulnery; cytostatic;
KM haemopoietic; haematologic; anaemia; autoimmune disorder;
KM rheumatoid arthritis; inflammation; Grave's disease; diabetes;
KM systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
KM Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
KM cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
human.

OS Homo sapiens.
XX MO2003038063-A2.
XX 08-MAY-2003.
XX 19-MAR-2002; 2002WO-US008277.
XX 21-MAR-2001; 2001US-0277340P.
XX 19-JUL-2001; 2001US-0306171P.
XX 13-NOV-2001; 2001US-0331287P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI: 2003-430516/40.
XX N-PSDB; ADC73950.
XX New human secreted polypeptide for diagnosing, preventing or treating
PT hematopoietic or hematologic disorders (e.g. anemia), autoimmune
PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or
PT atherosclerosis).

PS Claim 16; SEQ ID NO 1198; 2272pp; English.

XX The invention relates to a novel human secreted polypeptide comprising a
CC defined sequence given in the specification. The polypeptide, nucleic
CC acid molecule, antibody, agonist or antagonist of the invention may be
CC useful for preparing a composition for diagnosing or treating a
CC haemopoietic or haematologic disorder such as anaemia, autoimmune
CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,
CC diabetes, systemic lupus erythematosus or glomerulonephritis,
CC neurodegenerative disorders including Parkinson's disease and Alzheimer's
CC disease, wounds and hyperproliferative disorders including
CC atherosclerosis or cancer, as well as bacterial, viral, fungal or
CC parasitic infections. The polypeptide may also be used during gene
CC therapy procedures and for identifying a binding partner by contacting
CC the polypeptide with a binding partner and determining whether the
CC binding partner increases or decreases the activity of the polypeptide.
CC The current sequence is that of the human secreted protein of the
CC invention.

XX Sequence 305 AA;

SQ Query Match 100.0%; Score 1634; DB 7; Length 305;

Best Local Similarity 100.0%; Pred. No. 3.2e-161; Indels 0; Gaps 0;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLILGLSAGGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLILGLSAGGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PHLRLSGKCSLVSTYKYEFCPPHNTQHEQTFRRNAYSGIIGIHEWEIANNFTG 120
DB 61 PHLRLSGKCSLVSTYKYEFCPPHNTQHEQTFRRNAYSGIIGIHEWEIANNFTG 120
QY 121 MMWRDDACRSRSRQSKVELACGSKNRLAHVSEPTCYALTFETPLVCHPALVYPTL 180
DB 121 MMWRDDACRSRSRQSKVELACGSKNRLAHVSEPTCYALTFETPLVCHPALVYPTL 180
QY 181 PEALQROMQVEQDLADELITPQGHKELRLTFEDAGYLKTEENEPQLLEGSPSLGFE 240
DB 181 PEALQROMQVEQDLADELITPQGHKELRLTFEDAGYLKTEENEPQLLEGSPSLGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPRTSNLEHGHETPRAPKSPQQLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPRTSNLEHGHETPRAPKSPQQLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 11

ADD27815 ID ADD27815 standard; protein; 305 AA.

AC ADD27815;

DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase gamma subunit.

XX human; protein phosphorylation; soluble GlcNAc-phosphotransferase;

KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.

OS Homo sapiens.

XX US2003119088-A1.

XX 26-JUN-2003.

XX 21-DEC-2001; 2001US-00023888.

XX 21-DEC-2001; 2001US-00023888.

PA (NOVA-) NOVAZYME PHARM INC.

XX Canfield W, Kudo M;

XX WPI; 2003-801323/75.

DR N-PSDB; ADD27814.

XX Phosphorylating a protein for treating a patient suffering from a

PT lysosomal storage disease e.g. Fabry's disease by contacting the protein

PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated

PT protein.

XX Claim 11; SEQ ID NO 7; 55pp; English.

XX The invention relates to a method of phosphorylating a protein comprising

CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-

CC acetylglucosamine) and producing a phosphorylated protein. The method is

CC useful for treating a patient suffering from a lysosomal storage disease

CC e.g. Fabry's disease. The present sequence represents the amino acid

CC sequence of the human GlcNAc-phosphotransferase gamma subunit.

XX Sequence 305 AA;

SQ Query Match 100.0%; Score 1634; DB 7; Length 305;

Best Local Similarity 100.0%; Pred. No. 3.2e-161; Indels 0; Gaps 0;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLILGLSAGGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLILGLSAGGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
QY 61 PHLRLSGKCSLVSTYKYEFCPPHNTQHEQTFRRNAYSGIIGIHEWEIANNFTG 120
DB 61 PHLRLSGKCSLVSTYKYEFCPPHNTQHEQTFRRNAYSGIIGIHEWEIANNFTG 120
QY 121 MMWRDDACRSRSRQSKVELACGSKNRLAHVSEPTCYALTFETPLVCHPALVYPTL 180
DB 121 MMWRDDACRSRSRQSKVELACGSKNRLAHVSEPTCYALTFETPLVCHPALVYPTL 180
QY 181 PEALQROMQVEQDLADELITPQGHKELRLTFEDAGYLKTEENEPQLLEGSPSLGFE 240
DB 181 PEALQROMQVEQDLADELITPQGHKELRLTFEDAGYLKTEENEPQLLEGSPSLGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPRTSNLEHGHETPRAPKSPQQLRGDPG 300
DB 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPRTSNLEHGHETPRAPKSPQQLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 12

ABM01490 ID ABM01490 standard; protein; 305 AA.

AC ABM01490;

DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase gamma subunit protein.

XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell;

KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;

KW gastrointestinal; human; enzyme; lectin resistant cell;

KW deoxymannojirimycin; kifunensine; glycosylation inhibition.

XX Homo sapiens.

XX Key

FT Peptide

FT Protein

Location/Qualifiers

1..24

/label= Signal_peptide

25..305

FT		/note= "Mature human GlcNAc-phosphotransferase gamma
FT		subunit protein"
FN	US2003124652-A1.	
PD	03-JUL-2003.	
PF	21-DEC-2001; 2001US-00023889.	
PR	21-DEC-2001; 2001US-00023889.	
XX	(NOVA-) NOVAZYME PHARM INC.	
PA		
PI	Canfield WM;	
DR	WPI; 2003-810984/76.	
N-PSDB;	AAD62492.	
PT	Producing a high mannose glycoprotein for treating lysosomal storage diseases, comprises culturing the lectin resistant mammalian cell in the presence of deoxymannojirimycin and kifunensine.	
ES	Claim 9; Page 20; 46pp; English.	
XX	The invention relates to a method for producing a high mannose glycoprotein. The method comprises: introducing and expressing a polynucleotide encoding a glycoprotein into a mammalian cell; culturing the cell in the presence of a lectin to obtain a lectin resistant cell; isolating the cell; culturing the cell in the presence of deoxymannojirimycin and kifunensine to inhibit glycosylation of the glycoprotein; and collecting the glycoprotein. The invention is useful in gene therapy. The method is useful for producing a high mannose glycoprotein in a complex carbohydrate deficient cell for treating lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase gamma subunit protein	
XX	Sequence 305 AA;	
SQ		
Query Match	100.0%; Score 1634; DB 7; Length 305;	
Best Local Similarity	100.0%; Pred.No. 3,2e-161;	
Matches 305; Conservative	0; Mismatches 0; Indels 0; Gaps 0	
QY	1 MAAGIARILLILIGLSAGGPAPRGAAKMKVVEPNAPGVNPPPLQAASRLQARDSPSVSG	60
DB	1 MAAGIARILLILIGLSAGGPAPRGAAKMKVVEPNAPGVNPPPLQAASRLQARDSPSVSG	60
QY	PVHLFRLSGKCSLVESTKYKFCFPFHNTQHEOFFRNAYSGIIIMHETIANNTFG	120
DB	61 PVHLFRLSGKCSLVESTKYKFCFPFHNTQHEOFFRNAYSGIIIMHETIANNTFG	120
QY	121 MMNRDGDACRSRSRSQKVYLACGSKNRLAHVSEPSTCVYALTFETPLVCPHALLVPTL	180
DB	121 MMNRDGDACRSRSRSQKVYLACGSKNRLAHVSEPSTCVYALTFETPLVCPHALLVPTL	180
QY	181 PEALOROMDOVQODLADELITPOGHEKLRTLTFEAGYIKTPENNEPQLBGGPDSLGE	240
DB	181 PEALOROMDOVQODLADELITPOGHEKLRTLTFEAGYIKTPENNEPQLBGGPDSLGE	240
QY	241 TLENCRKAHKELSKIKLKGLTLTHGIPYTRPTSTNSLEHGHTPRAKSPEQRGDPG	300
DB	241 TLENCRKAHKELSKIKLKGLTLTHGIPYTRPTSTNSLEHGHTPRAKSPEQRGDPG	300
QY	301 LRGSLLRGSL	305
DB	301 LRGSLLRGSL	305
RESULT 13		
ID	ADD37973 standard; protein; 305 AA.	
XX	ADD37973;	
AC		

DT	15-JAN-2004 (first entry)	
XX	Human secreted protein #156.	
XX	human secreted protein; Anti-allergic; Anti-inflammatory; Antibacterial;	
XX	Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic.	
XX	Homo sapiens.	
XX	WO200290526-A2.	
XX	14-NOV-2002.	
XX	19-MAR-2002; 2002WO-US008279.	
XX	21-MAR-2001; 2001US-0277340P.	
XX	19-JUL-2001; 2001US-0306171P.	
XX	13-NOV-2001; 2001US-0331287P.	
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Rosen CA, Ruben SM;	
XX	WPI; 2003-140218/13.	
XX	New human secreted proteins and nucleic acid molecules, useful for	
XX	preparing a diagnostic or pharmaceutical composition for diagnosing or	
XX	treating allergic or asthmatic disorders, or related immediate	
XX	hypersensitivity disorders.	
XX	Claim 1; SEQ ID NO 455; 1323bp; English.	
XX	The present invention relates to an isolated polypeptide or human	
XX	secreted protein. The polypeptides, nucleic acid molecules, antibodies or	
XX	their fragments, and agonists or antagonists that bind are useful for	
XX	preparing a diagnostic or pharmaceutical composition for diagnosing or	
XX	treating allergic or asthmatic disorders. The polypeptide is also useful	
XX	for identifying a binding partner by contacting the polypeptide with a	
XX	binding partner, and determining whether the binding partner increases or	
XX	decreases the activity of the polypeptide. The polypeptides and nucleic	
XX	acid molecules are also useful for detecting, preventing, diagnosing,	
XX	prognosticating, treating or ameliorating inflammatory disorders	
XX	neoplastic diseases, wound healing and disorders of epithelial cell	
XX	proliferation, immune disorders, cardiovascular disorders, blood-related	
XX	disorders, infectious diseases, endocrine disorders, or gastrointestinal	
XX	disorders. The nucleic acids are also useful for chromosome	
XX	identification, radiation hybrid mapping or long-range restriction	
XX	mapping, as molecular weight markers, or as hybridization or diagnostic	
XX	probes. The polypeptides and antibodies are useful for providing	
XX	immunological probes for differential identification of the tissues	
XX	immunohistochemistry assays. The present sequence represents a human	
XX	secreted protein.	
XX	Sequence 305 AA:	
XX	Query Match 100.0%; Score 1634; DB 7; Length 305;	
XX	Best Local Similarity 100.0%; Pred. No. 3.2e-161; Indels 0; Gaps 0;	
XX	Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MAAGIARLLLLLGASGAPAPAGAAKKKVEEPAFGVNNPFLQASRLQAKRDPSPVSG 60	
DB	1 MRAQLALLLLLLIGSAGGPAPAGAAKKKVEEPAFGVNNPFLQASRLQAKRDPSPVSG 60	
QY	61 PVHLFRLSGKCSFVSEYTYKKEPCPPFNVTQHEOTFFRNAYSGILGTHWEIANNFTTG 120	
DB	61 PVHLFRLSGKCSFVSEYTYKKEPCPPFNVTQHEOTFFRNAYSGILGTHWEIANNFTTG 120	
QY	121 MMWMDGACSRSSQSVVELACGSKNNLAHVSESTCVYALTPTPLVCPHALLVPTL 180	
DB	121 MMWMDGACSRSSQSVVELACGSKNNLAHVSESTCVYALTPTPLVCPHALLVPTL 180	
QY	181 PEALQRMDQVEDDLABELITPQGHKLRLLTFPDAGYLLKTPESNEPTLEGSDSIFGE 240	

Db 181 PEALQRMQVQEDLADLELITPGHEKILRTLFEDAGYIKTPEENEPTOLEGGPDSLGFE 240
 QY 241 TLNCRKAHKEISKEIKRLKGLITQHGIPYTRPTETSNLEHGHETPRKSPPEQLRGDPG 300
 Db 241 TLNCRKAHKEISKEIKRLKGLITQHGIPYTRPTETSNLEHGHETPRKSPPEQLRGDPG 300
 QY 301 LRGS 305
 Db 301 LRGS 305

RESULT 14
 ADD38084
 ID ADD38084 standard; protein; 305 AA.

XX ADD38084;
 AC
 DT 15-JAN-2004 (first entry)
 XX
 DE Human secreted protein #267.
 XX
 KW human secreted protein; Antiallergic; Antiinflammatory; Antibacterial;
 KM Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic.
 OS Homo sapiens.
 XX
 PN MO200290526-A2.
 XX
 PD 14-NOV-2002.
 PF 19-MAR-2002; 2002WO-US008279.
 XX
 PR 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-140218/13.
 XX
 PT New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing or
 PT treating allergic or asthmatic disorders, or related immediate
 PT hypersensitivity disorders.
 XX
 PS Claim 1; SEQ ID NO 566; 1323pp; English.

XX The present invention relates to an isolated polypeptide or human
 CC secreted protein. The polypeptides, nucleic acid molecules, antibodies or
 CC their fragments, and agonists or antagonists that bind are useful for
 CC preparing a diagnostic or pharmaceutical composition for diagnosing or
 CC treating allergic or asthmatic disorders. The polypeptide is also useful
 CC for identifying a binding partner by contacting the polypeptide with a
 CC binding partner, and determining whether the binding partner increases or
 CC decreases the activity of the polypeptide. The polypeptides and nucleic
 CC acid molecules are also useful for detecting, preventing, diagnosing,
 CC prognosticating, treating or ameliorating inflammatory disorders
 CC neoplastic diseases, wound healing and disorders of epithelial cell
 CC proliferation, immune disorders, cardiovascular disorders, blood-related
 CC disorders, infectious diseases, endocrine disorders, or gastrointestinal
 CC disorders. The nucleic acids are also useful for chromosome
 CC identification, radiation hybrid mapping or long-range restriction
 CC mapping, as molecular weight markers, or as hybridization or diagnostic
 CC probes. The polypeptides and antibodies are useful for providing
 CC immunological probes for differential identification of the tissues
 CC immunohistochemistry assays. The present sequence represents a human
 CC secreted protein.
 XX

SO Sequence 305 AA;

Query Match 100.0%, Score 1634, DB 7, Length 305;

Best Local Similarity 100.0%; Pred. No. 3.2e-161;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAGLARLLILLGLSAGGAPAGAAKVVESPNAGVNNPFLPOASRLQAKRDPSPVG 60
 Db 1 MAAGLARLLILLGLSAGGAPAGAAKVVESPNAGVNNPFLPOASRLQAKRDPSPVG 60
 QY 61 PVHLFRLSGKCFSLVSTYKYEFCPPHNTQHQEFRRMAVSGILGIMHEWLIANTFTG 120
 Db 61 PVHLFRLSGKCFSLVSTYKYEFCPPHNTQHQEFRRMAVSGILGIMHEWLIANTFTG 120
 QY 121 MMRDGDACRSRSRQSVELACGSNRLAHVSEPTCVYALTFETPLVCHPALVYPTL 180
 Db 121 MMRDGDACRSRSRQSVELACGSNRLAHVSEPTCVYALTFETPLVCHPALVYPTL 180
 QY 181 PEALQRMQVQEDLADLELITPGHEKILRTLFEDAGYIKTPEENEPTOLEGGPDSLGFE 240
 Db 181 PEALQRMQVQEDLADLELITPGHEKILRTLFEDAGYIKTPEENEPTOLEGGPDSLGFE 240
 QY 241 TLNCRKAHKEISKEIKRLKGLITQHGIPYTRPTETSNLEHGHETPRKSPPEQLRGDPG 300
 Db 241 TLNCRKAHKEISKEIKRLKGLITQHGIPYTRPTETSNLEHGHETPRKSPPEQLRGDPG 300
 QY 301 LRGS 305
 Db 301 LRGS 305

RESULT 15
 AEM01539
 ID AEM01539 standard; protein; 305 AA.

XX AEM01539;
 AC
 DT 15-JAN-2004 (first entry)
 XX
 DE Human GlcNAc-phosphotransferase gamma subunit protein.
 XX
 KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;
 KM N-acetylglucosamine-1-phosphotransferase; gene therapy; human; enzyme.
 OS Homo sapiens.
 XX
 PI Key Location/Qualifiers
 FH Peptide 1..24
 FT Protein /label=Signal_peptide
 FT 25..305
 FT /note="Mature human GlcNAc-phosphotransferase gamma
 subunit protein"

XX US2003124653-A1.
 XX 03-JUL-2003.
 PD
 PF 21-DEC-2001; 2001US-00023890.
 PR 21-DEC-2001; 2001US-00023890.
 PA (NOVA-) NOVAZYME PHARM INC.
 XX
 PI Canfield WM;
 XX
 DR WPI; 2003-810985/76.
 DR N-PSDB; AAD62651.
 XX
 PT Producing a glycoprotein with reduced complex carbohydrates by culturing
 PT the lectin resistant mammalian cell expressing the glycoprotein for
 PT treating lysosomal storage disease.
 XX
 PS Claim 9; Page 20; 46pp; English.

XX The present invention provides a method of producing a glycoprotein
 CC having reduced complex carbohydrates by culturing the lectin resistant

CC mammalian cell expressing the glycoprotein. The method is useful for
 CC producing a glycoprotein with reduced complex carbohydrates for treating
 CC lysosomal storage disease. The present invention is also useful in gene
 CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-
 CC phosphotransferase gamma subunit precursor cDNA
 XX

SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; length 305;
 Best Local Similarity 100.0%; Pred. No. 3.2e-161;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAGIARILLLLIGSAGGPAPAGAKMKVVEEPNAPGVNMPFLPGASRLQAKRDPSPVSG	60
DB	1	MAAGIARILLLLIGSAGGPAPAGAKMKVVEEPNAPGVNMPFLPGASRLQAKRDPSPVSG	60
QY	61	PVHLFRLSGKCFSLVESTYKYEFCEPHNVTOHEQTFRNNAYSGLIGIWHMEIANNTFTG	120
DB	61	PVHLFRLSGKCFSLVESTYKYEFCEPHNVTOHEQTFRNNAYSGLIGIWHMEIANNTFTG	120
QY	121	MMMRDGDACRSRROSQKVELACGKSNRLAHVSEBSTCYVALTFETPLVCHPHALLVPTL	180
DB	121	MMMRDGDACRSRROSQKVELACGKSNRLAHVSEBSTCYVALTFETPLVCHPHALLVPTL	180
QY	181	PEALORWDQVEQDLADELITPQHEKILRTLFEDAGYLKTPENEPPTOLEGSPDSLGE	240
DB	181	PEALORWDQVEQDLADELITPQHEKILRTLFEDAGYLKTPENEPPTOLEGSPDSLGE	240
QY	241	TLNCRKAHKEISKEIKELKGLLTQHGIPYTRPETSNIHLGHETPRAKSPBQLRGDPG	300
DB	241	TLNCRKAHKEISKEIKELKGLLTQHGIPYTRPETSNIHLGHETPRAKSPBQLRGDPG	300
QY	301	LRGSL	305
DB	301	LRGSL	305

Search completed: July 26, 2004, 11:09:12
 Job time : 30.138 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:06:34 ; Search time 8.59705 Seconds

(without alignments)
1831.549 Million cell updates/sec

Title: US-10-657-280-3

Perfect score: 1634

Sequence: 1 MAAGLARIILLLGLSAGGPA.....TPRAKSPQRLGDPGLRSL 305

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

- 1: Issued Patents AA:*
- 2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1634	100.0	305	4	US-09-635-872A-3
2	1634	100.0	305	4	US-09-636-077A-3
3	1634	100.0	305	4	US-09-636-060C-3
4	1634	100.0	305	4	US-09-986-552-3
5	1288.5	78.9	307	4	US-09-635-872A-9
6	1288.5	78.9	307	4	US-09-636-077A-9
7	1288.5	78.9	307	4	US-09-636-060C-9
8	1288.5	78.9	307	4	US-09-986-552-9
9	137.5	8.4	109	4	US-09-437-054A-16
10	119.5	7.3	83	4	US-09-437-054A-14
11	104	6.4	21	4	US-09-635-872A-28
12	104	6.4	21	4	US-09-636-077A-28
13	104	6.4	21	4	US-09-636-060C-28
14	104	6.4	21	4	US-09-986-552-28
15	93.5	5.7	378	4	US-09-199-637A-167
16	93.5	5.7	419	4	US-09-252-991A-21423
17	92.5	5.7	216	1	US-07-917-722-4
18	91.5	5.6	216	2	US-08-489-666C-5
19	91.5	5.6	216	3	US-09-057-963A-4
20	91.5	5.6	1014	4	US-09-936-724-2
21	90	5.5	2584	4	US-08-936-135-4
22	88	5.4	593	2	US-08-900-927-4
23	88	5.4	593	2	US-09-191-279-4
24	88	5.4	593	3	US-09-334-476-4
25	88	5.4	935	4	US-09-314-258-25
26	86	5.3	192	4	US-09-107-532A-6267
27	84.5	5.2	398	4	US-08-406-824A-8

28	83	5.1	529	4	US-09-252-991A-29757	Sequence 29757, A
29	83	5.1	635	4	US-09-489-039A-10374	Sequence 10374, A
30	82.5	5.0	711	4	US-09-489-039A-12188	Sequence 12188, A
31	81.5	5.0	360	4	US-09-540-236-2029	Sequence 2029, Ap
32	81.5	5.0	422	3	US-09-109-204-30	Sequence 30, Appl
33	81.5	5.0	422	4	US-09-490-032-30	Sequence 30, Appl
34	81.5	5.0	819	4	US-09-976-594-369	Sequence 369, App
35	81.5	5.0	1112	4	US-09-252-991A-27256	Sequence 27256, A
36	81	5.0	744	4	US-09-252-991A-19290	Sequence 19290, A
37	80.5	4.9	292	4	US-09-724-623-81	Sequence 81, Appl
38	80	4.9	249	1	US-08-450-257-57	Sequence 57, Appl
39	80	4.9	249	1	US-08-450-246-57	Sequence 57, Appl
40	80	4.9	249	1	US-08-450-098-57	Sequence 57, Appl
41	80	4.9	249	1	US-08-451-233-57	Sequence 57, Appl
42	80	4.9	249	1	US-08-450-236-57	Sequence 57, Appl
43	80	4.9	249	4	US-08-235-403-57	Sequence 57, Appl
44	80	4.9	312	1	US-08-094-128A-27	Sequence 27, Appl
45	80	4.9	312	1	US-08-455-674-27	Sequence 27, Appl

ALIGNMENTS

```
RESULT 1
US-09-635-872A-3
; Sequence 3, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OR INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
US-09-635-872A-3
```

Query Match 100.0%; Score 1634; DB 4; Length 305;

Best Local Similarity 100.0%; Pred. No. 5.6e-178; Mismatches 0; Indels 0; Gaps 0;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAGLARIILLLGLSAGGPAAGAAKVEEENAFVNNPFLPQASRLQAKRDPSPVSG	60
DB	1	MAAGLARIILLLGLSAGGPAAGAAKVEEENAFVNNPFLPQASRLQAKRDPSPVSG	60
QY	61	PVHLFRLSGKCFELVSTYKVEFCPEFHYVTOHQETRRMAVSGIIGIMHEWELANTFTG	120
DB	61	PVHLFRLSGKCFELVSTYKVEFCPEFHYVTOHQETRRMAVSGIIGIMHEWELANTFTG	120
QY	121	MMNRDADGSRSGKQKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL	180
DB	121	MMNRDADGSRSGKQKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTL	180
QY	181	PEALQOMQVQVEDLDELITPGGHEKILRTLFEDAGYIKTPEENPTOLGGPDSLGRE	240
DB	181	PEALQOMQVQVEDLDELITPGGHEKILRTLFEDAGYIKTPEENPTOLGGPDSLGRE	240
QY	241	TLENCKRAHEKSEIKRLKGLITTOGIPYRPTETSNLEHGHETPRAKSPFOALRGDGG	300
DB	241	TLENCKRAHEKSEIKRLKGLITTOGIPYRPTETSNLEHGHETPRAKSPFOALRGDGG	300
QY	301	LRGSL 305	
DB	301	LRGSL 305	

```
RESULT 2
US-09-636-077A-3
; Sequence 3, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
US-09-636-077A-3

Query Match          100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 5.6e-178;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
      |||
      1  MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB
QY      61  PVHLFRLSGKCSFLVSTYKYEFCEPHNVTQHEQTFRNNAYSGIIGIWEWEIANNFTFG 120
      |||
      61  PVHLFRLSGKCSFLVSTYKYEFCEPHNVTQHEQTFRNNAYSGIIGIWEWEIANNFTFG 120
DB
QY      121  MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
      |||
      121  MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
DB
QY      181  PEALQRMQVQEDLADDELITPQGHKRLRTLFEDAGYLKTPENEPTQLEGGPDLSGFE 240
      |||
      181  PEALQRMQVQEDLADDELITPQGHKRLRTLFEDAGYLKTPENEPTQLEGGPDLSGFE 240
DB
QY      241  TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
      |||
      241  TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
DB
QY      301  LRGSLL 305
      |||
      301  LRGSLL 305
DB
301  LRGSLL 305

RESULT 3
US-09-636-060C-3
; Sequence 3, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US00CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-09-636-060C-3

Query Match          100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 5.6e-178;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
      |||
      1  MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB
QY      61  PVHLFRLSGKCSFLVSTYKYEFCEPHNVTQHEQTFRNNAYSGIIGIWEWEIANNFTFG 120
      |||
      61  PVHLFRLSGKCSFLVSTYKYEFCEPHNVTQHEQTFRNNAYSGIIGIWEWEIANNFTFG 120
DB
QY      121  MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
      |||
      121  MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
DB
QY      181  PEALQRMQVQEDLADDELITPQGHKRLRTLFEDAGYLKTPENEPTQLEGGPDLSGFE 240
      |||
      181  PEALQRMQVQEDLADDELITPQGHKRLRTLFEDAGYLKTPENEPTQLEGGPDLSGFE 240
DB
QY      241  TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
      |||
      241  TLNCRKAHKEISKEIKRLKGLLTQHGIPYTPPTETSNLEHGHETPRAKSPDQLRGDPG 300
DB
QY      301  LRGSLL 305
      |||
      301  LRGSLL 305
DB
301  LRGSLL 305

RESULT 4
US-09-986-552-3
; Sequence 3, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-09-986-552-3

Query Match          100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 5.6e-178;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
      |||
      1  MAAGLARLLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPOASRLQAKRDPSPVSG 60
DB
QY      61  PVHLFRLSGKCSFLVSTYKYEFCEPHNVTQHEQTFRNNAYSGIIGIWEWEIANNFTFG 120
      |||
      61  PVHLFRLSGKCSFLVSTYKYEFCEPHNVTQHEQTFRNNAYSGIIGIWEWEIANNFTFG 120
DB
QY      121  MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
      |||
      121  MMRRDDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPALLVYPTL 180
```

```

Db      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPVLVCHPHALLVPTL 180
Qy      181 PEARLORQDQVODLADDELITPOGHEKILRLTFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Db      181 PEARLORQDQVODLADDELITPOGHEKILRLTFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Qy      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPAKSPEDQLRGDPG 300
Db      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPAKSPEDQLRGDPG 300
Qy      301 LRGS 305
Db      301 LRGS 305

```

RESULT 5

```

US-09-635-872A-9
; Sequence 9, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-635-872A-9.

```

Query Match 78.9%; Score 1288.5; DB 4; Length 307;

Best Local Similarity 78.4%; Pred. No. 1.7e-138;

Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

```

Qy      1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
Db      1 MAGRLAGFLMLLGLASQGPAPACAGKMKVVEBPNTFGLNPFLLPQASRLQKREPSAVSG 60
Qy      61 PVLHRLSGKCFSLVSTYKYECPPHANTQHEQTFRRNNAVSGIIGIWHMEIANTPTG 120
Db      61 PVLHRLAGKCFSLVSTYKYECPPHANTQHEQTFRRNNAVSGIIGIWHMEIANTPTG 120
Qy      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPVLVCHPHALLVPTL 180
Db      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPVLVCHPHALLVPTL 180
Qy      181 PEARLORQDQVODLADDELITPOGHEKILRLTFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Db      181 PEARLORQDQVODLADDELITPOGHEKILRLTFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Qy      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPAKSPEDQLRGDP 299
Db      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPAKSPEDQLRGDP 299
Qy      300 LRGS 305
Db      301 LRGS 306

```

RESULT 6

```

US-09-636-077A-9
; Sequence 9, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0

```

```

; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-077A-9

```

Query Match 78.9%; Score 1288.5; DB 4; Length 307;

Best Local Similarity 78.4%; Pred. No. 1.7e-138;

Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

```

Qy      1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
Db      1 MAGRLAGFLMLLGLASQGPAPACAGKMKVVEBPNTFGLNPFLLPQASRLQKREPSAVSG 60
Qy      61 PVLHRLSGKCFSLVSTYKYECPPHANTQHEQTFRRNNAVSGIIGIWHMEIANTPTG 120
Db      61 PVLHRLAGKCFSLVSTYKYECPPHANTQHEQTFRRNNAVSGIIGIWHMEIANTPTG 120
Qy      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPVLVCHPHALLVPTL 180
Db      121 MMRDGDACRSRSGKVELACGKSNRLAHVSEPTCYVALTFEPVLVCHPHALLVPTL 180
Qy      181 PEARLORQDQVODLADDELITPOGHEKILRLTFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Db      181 PEARLORQDQVODLADDELITPOGHEKILRLTFEDAGYLKTPENEPPTOLGSPDSIGFE 240
Qy      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPAKSPEDQLRGDP 299
Db      241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPETSULEHGHETPAKSPEDQLRGDP 299
Qy      300 LRGS 305
Db      301 LRGS 306

```

RESULT 7

```

US-09-636-060C-9
; Sequence 9, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210419US00CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-9

```

Query Match 78.9%; Score 1288.5; DB 4; Length 307;

Best Local Similarity 78.4%; Pred. No. 1.7e-138;

Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

```

Qy      1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEBPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
Db      1 MAGRLAGFLMLLGLASQGPAPACAGKMKVVEBPNTFGLNPFLLPQASRLQKREPSAVSG 60
Qy      61 PVLHRLSGKCFSLVSTYKYECPPHANTQHEQTFRRNNAVSGIIGIWHMEIANTPTG 120
Db      61 PVLHRLAGKCFSLVSTYKYECPPHANTQHEQTFRRNNAVSGIIGIWHMEIANTPTG 120

```

Qy	12	MMNRDGCAGSRSSRSQKVELACGSKSRILAHVSEPSVCYALTFEETPLVCHPHALLVPTL	180
	121	MMMTDGDSCSRSSRSQKVELTCGKINRLAHVSEPSVCYALTFEETPLVCHPHALLVPTL	180
Qy	181	PEALQRCMDVEODLADDELTTPQGHKILRTIFEDAGYLKTPREENPPTQEGSPDSAGFE	240
	181	SEALQRCMDVEODLADDELTTPQGYKILRLVFEFAGYLKTPGHEHNPQLQAGSGKGLGE	240
Qy	241	TLENCRKAHKELSKETIKRLKGLLTHGILPYTPETLSNLEHGHETPRAKSPQLRGDP	299
	241	TLENCRKAHAELEQEVORLTSLQGHGILPHTPTETTSOHLGQCLPGIAYAAHLESDP	300
Qy	300	GLRGSL 305	
	301	GLRGNI 306	

```

RESULT 8
US-09-986-552-9
; Sequence 9, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215083U77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-552-9

```

	Query Match	78.9%	Score 1288.5	DB 4	Length 307
	Best Local Similarity	78.4%	Pred. No. 1.7e-118		
	Matches 240	Conservative 25	Mismatches 40	Indels 1	Gaps 1
QY	1	MAAGLRLILLIGLSGAPAPAGAAMKVEBPNAFGVNNPFLPQASRLQAKEDSPSPVG	60		
Db	1	MAARLGLPMLLIGLASQGPAPACAGMKVVEBPNTFGANPFLPQASRLQPKREPAVSG	60		
QY	61	PVFLPFLSGKCSLVESTYKFEFCPFNNVTHQSDTFPMNAYSGILGIWHEWBIANTFFG	120		
Db	61	PLFLPFLAKCSLVESTYKFEFCPFNNVTHQSDTFPMNAYSGILGIWHEWBIINNTFFG	120		
QY	121	MMNRDDADGRSSRQSKVELACGKSNRLAHSEPSSTCVALLTEPTLVCHPALLLVPTL	180		
Db	121	MMNTDDSDSHSSRSQSKVELTCGKNNRLAHVSEPSSTCVALLTEPTLVCHPSLLVPTL	180		
QY	181	PEALQROMQOVQCDLADDELITPQGHKKLRLTFPDAGLYTPPEENEPQLGGSDLSGFE	240		
Db	181	SEMLQGRRLQOVQCDLADDELITPQGYEKKLRLVFDAGLYKVPETHTPQLAGSKSLGFE	240		
QY	241	TLENCKRAHELSKEIKRLKGLLTHQGIPTYRPTETSNLEHLGHETP-RAKSPDQLGDP	299		
Db	241	TLDNCKRAAEHSQEVQRILTSILQHGHIPTQPIETHSQHLGQQLPQIGALIAEHLSRDP	300		
QY	300	GLRGSL	305		
Db	301	GLRGNT	306		

```

1  APPLICANT: Allen, Stephen M.
2  APPLICANT: Kinney, Anthony J.
3  TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
4  FILE REFERENCE: BB1273 US NA
5  CURRENT APPLICATION NUMBER: US/09/437,054A
6  CURRENT FILING DATE: 2001-05-14
7  PRIOR APPLICATION NUMBER: 60/107,909
8  PRIORITY FILING DATE: 1998-No. 6316698ember-10
9  NUMBER OF SEQ. ID NOS: 19
10 SOFTWARE: Microsoft Office 97
11 SEQ. ID NO. 16
12 LENGTH: 109
13 TYPE: PR1
14 ORGANISM: Triticum aestivum
15 US-09-437-054A-16

```

```

Query Match      8 4% Score 137.5; DB 4; Length 109;
Best Local Similarity 31.7%; Pred. No. 1e-07;
Matches 40; Conservative 13; Mismatches 50; Indels 23; Gaps 5

QY      CPFHNVTOHEQTFRRNNAYSGILGIWHEMEIANTFTGMWRDGDAC-RRSRSQSRYELAC 142
        |||:::||||::|
Db       6 CFFKASQVEGHSSNT-----LGRWDKE--ESCRMHFNSNDCKMGNPDBSLKYRLRC 57

QY      GASNRIATYSSESTGVVLTETETPLVCHPHALLVTPIIPBALORQMDQVEQLADELTTP 202
        |||||:::|||||
Db       58 GLSNELNGVDSPSRCEVAVLSTAMCVEEKL-----KELQOKUDAASSPLS----- 104

QY      203 QGHEKL 208
        |||::|
Db       105 -GHDEL 109

```

```

RESULT 10
US-09-437-054A-14
Sequence 14, Application US/09437054A
Patent No. 6316698
GENERAL INFORMATION:
APPLICANT: Allen, Stephen M.
APPLICANT: Kinney, Anthony J.
TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
FILE REFERENCE: B11573 US NA
CURRENT APPLICATION NUMBER: US/09/437,054A
PRIORITY FILING DATE: 2001-05-14
PRIOR APPLICATION NUMBER: 60/107,909
PRIORITY FILING DATE: 1999-No. 6316698ember-10
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Microsoft Office 97
SEQ ID NO 14
LENGTH: 83
TYPE: PRT
ORGANISM: Trillium aestivum
US-09-437-054A-14

```

Query Match	7.3%	Score 119.5	DB 4	Length 83
Best Local Similarity	35.6%	Pred. No. 7.5e-06		
Matches	31	Conservative	7	Mismatches 40; Indels 9; Gaps 3.

QY	84	CPEHNVQGEQIFRNAAVSGILGIMHEMLANNTTCGMMDQDAC-RSRQSQVYELAC	142
		
Db	3	CPYKAAQVEGHSSIN-----LGRWDKFE---ESCRMHFSNDDCKMNPDPASLKYRLRC	54
		
OY	143	GKSNRIAYSEPSICVYALTPETPLVC	169
		
Db	55	GLSNEIANGVDEPSRCEYVAIVSTPAMC	81
		

```

RESULT 9
US-09-437-054A-16
; Sequence 16, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:

```

RESULT 11
US-09-635-872A-28
; Sequence 28, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM

;; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
;; FILE REFERENCE: 195613US0
;; CURRENT APPLICATION NUMBER: US/09/635,872A
;; CURRENT FILING DATE: 2000-08-10
;; PRIOR APPLICATION NUMBER: 60/153,831
;; PRIOR FILING DATE: 1999-09-14
;; NUMBER OF SEQ ID NOS: 52
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 28
;; LENGTH: 21
;; TYPE: PRT
;; ORGANISM: Bos taurus
US-09-635-872A-28

Query Match 6.4%; Score 104; DB 4; Length 21;
Best Local Similarity 90.5%; Pred. No. 4.8e-05;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 25 AKMKVVEEPNAGVNNPFLPQ 45
Db 1 AKMKVVEEPNAGVNNPFLPQ 21

RESULT 12
US-09-636-077A-28
; Sequence 28, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-636-077A-28

Query Match 6.4%; Score 104; DB 4; Length 21;
Best Local Similarity 90.5%; Pred. No. 4.8e-05;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 25 AKMKVVEEPNAGVNNPFLPQ 45
Db 1 AKMKVVEEPNAGVNNPFLPQ 21

RESULT 13
US-09-636-060C-28
; Sequence 28, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 210119US00CT
; CURRENT APPLICATION NUMBER: US/09/636,060C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-636-060C-28

Query Match 6.4%; Score 104; DB 4; Length 21;

Best Local Similarity 90.5%; Pred. No. 4.8e-05;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 25 AKMKVVEEPNAGVNNPFLPQ 45
Db 1 AKMKVVEEPNAGVNNPFLPQ 21

RESULT 14
US-09-986-552-28
; Sequence 28, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-986-552-28

Query Match 6.4%; Score 104; DB 4; Length 21;
Best Local Similarity 90.5%; Pred. No. 4.8e-05;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 25 AKMKVVEEPNAGVNNPFLPQ 45
Db 1 AKMKVVEEPNAGVNNPFLPQ 21

RESULT 15
US-09-199-637A-367
; Sequence 367, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 367
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-367

Query Match 5.7%; Score 93.5; DB 4; Length 378;
Best Local Similarity 20.8%; Pred. No. 0.082; 76; Indels 113; Gaps 11;
Matches 57; Conservative 28; Mismatches 76; Indels 113; Gaps 11;

QY 23 GAAMKVEEPNAGVNNPFLPQASRLQAKRDPSPVSGFVHLFRLSGKCFSLVSTYKYE 82

```

Db      145 GAGFORAVYMSGVAVNPNRLARSLRASIQQ----- 175
Qy      83 FCFPHAVTQHEQTFPRNNAYSGILGIWHEWEIANTFTGMWEDGP-----ACRSRSRQS 136
Db      176 ---FANLELHEQT-----EVRG-----WLRDGRVVGVAATSRGEIRGD 210
Qy      137 KYELACGK-SNRLAHVSEPTCVALTFTETPLVCHPHALLVPTLPEALCRQMDQVEODL 195
Db      211 KYLLAAGAMSGEL-----LKPLGLELPVPVVKGMILYKCAADFLPR----- 252
Qy      196 ADELITPOGH-----EKLRITLPEDAGYIKTPENEPFTQLBGCPDLSGFTLENCRK 247
Db      253 ---MYLAKGRVAIPRRDGHILIGSTLEHSGFDXPTD-----EALLESIRA 294
Qy      248 AHKELSKKIRLK-----GLL--TOHGIPYTRP 273
Db      295 SAAELIPELADMQPVAMAGLRPGSPGIPYIGP 328

```

Search completed: July 26, 2004, 11:13:20
 Job time : 9.59705 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:11:44 ; Search time 23.4465 Seconds

(without alignments)
4073.614 Million cell updates/sec

Title: US-10-657-280-3

Perfect score: 1634

Sequence: 1 MAAGLARLLILLIGLSAGGPA.....TPRAKSPQKRGDPGLGSL 305

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 33154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1634	100.0	305	9	US-09-895-072-3
2	1634	100.0	305	9	US-09-986-552-3
3	1634	100.0	305	14	US-10-023-888-7
4	1634	100.0	305	14	US-10-023-889-7
5	1634	100.0	305	14	US-10-023-890-7
6	1634	100.0	305	14	US-10-024-197-7
7	1634	100.0	305	14	US-10-023-894-7
8	1634	100.0	305	14	US-10-306-886-3
9	1634	100.0	306	12	US-10-372-876-152
10	1634	100.0	306	14	US-10-097-065-152
11	1634	100.0	310	12	US-10-296-115-1319
12	1288.5	78.9	307	9	US-09-895-072-9
13	1288.5	78.9	307	9	US-09-986-552-9
14	1288.5	78.9	307	14	US-10-023-888-12
15	1288.5	78.9	307	14	US-10-023-889-12

15	1288.5	78.9	307	14	US-10-023-890-12	Sequence 12, Appl
17	1288.5	78.9	307	14	US-10-024-197-12	Sequence 12, Appl
18	1288.5	78.9	307	14	US-10-023-894-12	Sequence 12, Appl
19	1288.5	78.9	307	14	US-10-306-686-9	Sequence 9, Appl
20	1244.5	76.2	296	15	US-10-131-487A-112	Sequence 112, App
21	237	14.5	44	14	US-10-029-366-31490	Sequence 31490, A
22	205	12.5	38	14	US-10-029-366-31804	Sequence 31804, A
23	170	10.4	478	12	US-10-424-599-276459	Sequence 276459, A
24	169.5	10.4	330	12	US-10-425-114-47678	Sequence 47678, A
25	167.5	10.3	432	12	US-10-425-114-476502	Sequence 276502, A
26	156.5	9.6	278	12	US-10-425-114-41458	Sequence 41458, A
27	156.5	9.6	610	12	US-10-425-114-53637	Sequence 53637, A
28	155.5	9.5	597	15	US-10-331-496A-19	Sequence 650, App
29	155.5	9.5	612	12	US-10-072-012-650	Sequence 650, App
30	155.5	9.5	650	12	US-10-072-012-244	Sequence 244, App
31	155.5	9.5	667	12	US-10-072-012-649	Sequence 649, App
32	154	9.4	614	16	US-10-437-963-111038	Sequence 111038, A
33	149	9.1	668	12	US-10-072-012-242	Sequence 242, App
34	140.5	8.6	524	12	US-10-221-278-212	Sequence 212, App
35	140.5	8.6	524	15	US-10-291-172-212	Sequence 212, App
36	140.5	8.6	527	12	US-10-221-278-211	Sequence 211, App
37	140.5	8.6	527	12	US-10-257-021-20	Sequence 20, Appl
38	140.5	8.6	527	15	US-10-291-172-211	Sequence 211, App
39	140.5	8.6	527	16	US-10-408-765A-484	Sequence 484, App
40	140.5	8.6	534	12	US-10-221-278-213	Sequence 213, App
41	140.5	8.6	534	15	US-10-291-172-213	Sequence 213, App
42	130	8.0	571	12	US-10-072-012-652	Sequence 652, App
43	125	7.6	483	16	US-10-408-765A-1836	Sequence 1836, App
44	120.5	7.4	483	10	US-09-796-753-44	Sequence 44, Appl
45	119	7.3	483	12	US-10-221-278-310	Sequence 310, App

ALIGNMENTS

RESULT 1
US-09-895-072-3
; Sequence 3, Application US/09895072
; Patent No. US2002025550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OR INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119USOCONT
; CURRENT APPLICATION NUMBER: US/09/895, 072
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153, 831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635, 872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-09-895-072-3

Query Match 100.0%; Score 1634; DB 9; Length 305;

Best Local Similarity 100.0%; Pred. No. 1.3e-156;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAGLARLLILLIGLSAGGPA	PAGAAKMKVVEEBNA	FGVNNPFLPQASRLQAKRDP	SPVSG 60
DB	1	MAAGLARLLILLIGLSAGGPA	PAGAAKMKVVEEBNA	GVNNPFLPQASRLQAKRDP	SPVSG 60
QY	61	PVTLFRLSGKCFSLVSTYK	YECFPHANTQ	QHQTRMAVSGILGIW	MEIANTPTFG 120
DB	61	PVTLFRLSGKCFSLVSTYK	YECFPHANTQ	QHQTRMAVSGILGIW	MEIANTPTFG 120

QY 121 MMRDGDACRSRSRQSKVELACGKSNRLAHVSEBPTCVYALTFETPLVCHPALLVPTL 180
DB 121 MMRDGDACRSRSRQSKVELACGKSNRLAHVSEBPTCVYALTFETPLVCHPALLVPTL 180
QY 181 PEALORQWQVQDLADELITPGHEKILRTLFEADAGYLKTEPEENPTQLEGSPDSLGE 240
DB 181 PEALORQWQVQDLADELITPGHEKILRTLFEADAGYLKTEPEENPTQLEGSPDSLGE 240
QY 241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPTETSNLEHGHETTPRAKSPQRLGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPTETSNLEHGHETTPRAKSPQRLGDPG 300
QY 301 LRGS 305
DB 301 LRGS 305

RESULT 2
US-09-986-552-3
; Sequence 3, Application US/09986552
; Patent No. US20020150981A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-09-986-552-3

Query Match 100.0%; Score 1634; DB 9; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVG 60
DB 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVG 60
QY 61 PVHLFRLSGKCFSLVESTYKYFCPPHNTQHEQTFRMNAYSGIIGIHWHEMIANNFTG 120
DB 61 PVHLFRLSGKCFSLVESTYKYFCPPHNTQHEQTFRMNAYSGIIGIHWHEMIANNFTG 120
QY 121 MMRDGDACRSRSRQSKVELACGKSNRLAHVSEBPTCVYALTFETPLVCHPALLVPTL 180
DB 121 MMRDGDACRSRSRQSKVELACGKSNRLAHVSEBPTCVYALTFETPLVCHPALLVPTL 180
QY 181 PEALORQWQVQDLADELITPGHEKILRTLFEADAGYLKTEPEENPTQLEGSPDSLGE 240
DB 181 PEALORQWQVQDLADELITPGHEKILRTLFEADAGYLKTEPEENPTQLEGSPDSLGE 240
QY 241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPTETSNLEHGHETTPRAKSPQRLGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPTETSNLEHGHETTPRAKSPQRLGDPG 300
QY 301 LRGS 305
DB 301 LRGS 305

RESULT 3

US-10-023-888-7
; Sequence 7, Application US/10023888
; Publication No. US2003011908A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent version 3.1
; SEQ ID NO 7
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-888-7

Query Match 100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVG 60
DB 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVG 60
QY 61 PVHLFRLSGKCFSLVESTYKYFCPPHNTQHEQTFRMNAYSGIIGIHWHEMIANNFTG 120
DB 61 PVHLFRLSGKCFSLVESTYKYFCPPHNTQHEQTFRMNAYSGIIGIHWHEMIANNFTG 120
QY 121 MMRDGDACRSRSRQSKVELACGKSNRLAHVSEBPTCVYALTFETPLVCHPALLVPTL 180
DB 121 MMRDGDACRSRSRQSKVELACGKSNRLAHVSEBPTCVYALTFETPLVCHPALLVPTL 180
QY 181 PEALORQWQVQDLADELITPGHEKILRTLFEADAGYLKTEPEENPTQLEGSPDSLGE 240
DB 181 PEALORQWQVQDLADELITPGHEKILRTLFEADAGYLKTEPEENPTQLEGSPDSLGE 240
QY 241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPTETSNLEHGHETTPRAKSPQRLGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLITQHGIPYTRPTETSNLEHGHETTPRAKSPQRLGDPG 300
QY 301 LRGS 305
DB 301 LRGS 305

RESULT 4
US-10-023-889-7
; Sequence 7, Application US/10023889
; Publication No. US20030124652A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBO
; FILE REFERENCE: 203512US77
; CURRENT APPLICATION NUMBER: US/10/023,889
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent version 3.1
; SEQ ID NO 7
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-889-7

Query Match 100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVG 60
DB 1 MAAGLARLLLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVG 60

```

QY 61 PVHLFRLSGKCSFSLVSTYKYEFCPFHNVTQHEQTFRNMAVSGILGIWHEWEIANNPTG 120
DB 61 PVHLFRLSGKCSFSLVSTYKYEFCPFHNVTQHEQTFRNMAVSGILGIWHEWEIANNPTG 120
QY 121 MMNRDGDACRSRSRSQSKVELACGKSNRLAHVSEPSSTCYVALTFETPLVCHPHALLVYPTL 180
DB 121 MMNRDGDACRSRSRSQSKVELACGKSNRLAHVSEPSSTCYVALTFETPLVCHPHALLVYPTL 180
QY 181 PEALORQMDQVQEDLADDELITPQGEKILRTLFEDAGYLKTPBEENPTQLGGPDSLGFE 240
DB 181 PEALORQMDQVQEDLADDELITPQGEKILRTLFEDAGYLKTPBEENPTQLGGPDSLGFE 240
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPETSNIJLHGHETPRAKSPBQLRGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPETSNIJLHGHETPRAKSPBQLRGDPG 300
QY 301 LRGS 305
DB 301 LRGS 305

RESULT 5
US-10-023-890-7
; Sequence 7, Application US/10023890
; Publication No. US20030124653A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYD
; FILE REFERENCE: 203510US77
; CURRENT APPLICATION NUMBER: US/10/023, 890
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-890-7

Query Match      100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS
; FILE REFERENCE: 209794US0
; CURRENT APPLICATION NUMBER: US/10/024, 197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-024-197-7

Query Match      100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
US-10-023-894-7
; Sequence 7, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023, 894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-894-7

Query Match      100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6
US-10-024-197-7
; Sequence 7, Application US/10024197
; Publication No. US20030133924A1

```

Db 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRMNAYSGIIGIWHMEIANNTFTG 120
QY 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
Db 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALQROMQVQEDLADDELITPOGHEKILRTLFEDAGYLKTPPEENEPQLLEGSPDSIGFE 240
Db 181 PEALQROMQVQEDLADDELITPOGHEKILRTLFEDAGYLKTPPEENEPQLLEGSPDSIGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHLGHETPRAXSPBQURGDPG 300
Db 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHLGHETPRAXSPBQURGDPG 300
QY 301 LRGS 305
Db 301 LRGS 305

RESULT 8
US-10-306-686-3
; Sequence 3; Application US/10306686
; Publication No. US20030148460A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANase OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 230397US7Ddiv
; CURRENT APPLICATION NUMBER: US/10/306,686
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: 09/636,596
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-08-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-10-306-686-3

Query Match 100.0%; Score 1634; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLILLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
Db 1 MAAGLARLLILLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRMNAYSGIIGIWHMEIANNTFTG 120
Db 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRMNAYSGIIGIWHMEIANNTFTG 120
QY 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
Db 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALQROMQVQEDLADDELITPOGHEKILRTLFEDAGYLKTPPEENEPQLLEGSPDSIGFE 240
Db 181 PEALQROMQVQEDLADDELITPOGHEKILRTLFEDAGYLKTPPEENEPQLLEGSPDSIGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHLGHETPRAXSPBQURGDPG 300
Db 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHLGHETPRAXSPBQURGDPG 300
QY 301 LRGS 305
Db 301 LRGS 305

RESULT 9
US-10-372-876-152
; Sequence 152; Application US/10372876
; Publication No. US20030204071A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021P1
; CURRENT APPLICATION NUMBER: US/10/372,876
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 09/334,595
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: PCT/US98/27059
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/070,923
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,007
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,057
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,006
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,369
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,367
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,169
; PRIOR FILING DATE: 1997-12-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 152
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (306)
; OTHER INFORMATION: Xaa equals stop translation
US-10-372-876-152

Query Match 100.0%; Score 1634; DB 12; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGLARLLILLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
Db 1 MAAGLARLLILLLGLSAGPAPAGAAKMKVVEEPNAGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRMNAYSGIIGIWHMEIANNTFTG 120
Db 61 PVHLFRLSGKCFSLVSTYKYEFCFPHNTQHEQTFRMNAYSGIIGIWHMEIANNTFTG 120
QY 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
Db 121 MMRDGDACRSRSRSGKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALQROMQVQEDLADDELITPOGHEKILRTLFEDAGYLKTPPEENEPQLLEGSPDSIGFE 240
Db 181 PEALQROMQVQEDLADDELITPOGHEKILRTLFEDAGYLKTPPEENEPQLLEGSPDSIGFE 240
QY 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHLGHETPRAXSPBQURGDPG 300
Db 241 TLNCRKAHKEISKEIKRLKGLLTQHGIPYTRPTETSNLEHLGHETPRAXSPBQURGDPG 300
QY 301 LRGS 305
Db 301 LRGS 305

```
RESULT 10
US-10-097-065-152
; Sequence 152, Application US/10097065
; Publication No. US20030055236A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021PI
; CURRENT APPLICATION NUMBER: US/10/097,065
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: PCT/US98/27059
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/070,923
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,007
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,057
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,006
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,369
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,367
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,169
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,053
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,064
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,054
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,008
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,365
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 152
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (306)
; OTHER INFORMATION: Xaa equals stop translation
US-10-097-065-152

Query Match      100.0%; Score 1634; DB 14; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLGLSAGGPAPAGAAKMKVVEBNPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
DB 1 MAAGLARLLLLGLSAGGPAPAGAAKMKVVEBNPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PVHLFRLSGKCFSLVSTYKYEFCPPHNVTQHEQTRNNAYSGILGIWHEWEIANNPTFG 120
DB 61 PVHLFRLSGKCFSLVSTYKYEFCPPHNVTQHEQTRNNAYSGILGIWHEWEIANNPTFG 120
QY 121 MMNRDGDACRSRQSKVELACGSKNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
DB 121 MMNRDGDACRSRQSKVELACGSKNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALQROWDQVEODLADDELITPQGHKLLRTLTFEDAGYLKTPRENEPTQLSGPDSLGE 240
DB 181 PEALQROWDQVEODLADDELITPQGHKLLRTLTFEDAGYLKTPRENEPTQLSGPDSLGE 240
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
```

```
DB 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
QY 301 LRGSLL 305
DB 301 LRGSLL 305

RESULT 11
US-10-296-115-1319
; Sequence 1319, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hysq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1319
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-1319

Query Match      100.0%; Score 1634; DB 12; Length 310;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLLLLGLSAGGPAPAGAAKMKVVEBNPNAFGVNNPFLPQASRLQAKRDPSPVSG 60
DB 6 MAAGLARLLLLGLSAGGPAPAGAAKMKVVEBNPNAFGVNNPFLPQASRLQAKRDPSPVSG 65
QY 61 PVHLFRLSGKCFSLVSTYKYEFCPPHNVTQHEQTRNNAYSGILGIWHEWEIANNPTFG 120
DB 66 PVHLFRLSGKCFSLVSTYKYEFCPPHNVTQHEQTRNNAYSGILGIWHEWEIANNPTFG 125
QY 121 MMNRDGDACRSRQSKVELACGSKNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
DB 126 MMNRDGDACRSRQSKVELACGSKNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 185
QY 181 PEALQROWDQVEODLADDELITPQGHKLLRTLTFEDAGYLKTPRENEPTQLSGPDSLGE 240
DB 186 PEALQROWDQVEODLADDELITPQGHKLLRTLTFEDAGYLKTPRENEPTQLSGPDSLGE 245
QY 241 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
DB 246 TLNCRKAHKELSKEIKRLKGLLTQHGIPYTRPTETSNLEHGHETPRAKSPBQLRGDPG 305
QY 301 LRGSLL 305
DB 306 LRGSLL 310

RESULT 12
US-09-895-072-9
; Sequence 9, Application US/09895072
; Patent No. US2002002550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119US00CONT
; CURRENT APPLICATION NUMBER: US/09/895,072
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/635,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent In version 3.1
```

```

; SEQ ID NO 9
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-895-072-9

```

Query Match	78.9%;	Score 1288.5;	DB 9;	Length 307;
Best Local Similarity	78.4%;	Pred. No. 1.3e-121;		
Matches 240; Conservative	25;	Mismatches 40;	Indels 1;	Gaps 1.

QY 1 MAAGLARLLLLIGLSAGSPAPACAAKMKVEEPMAGVUNPFLPQASRLQAKRDEPSVG 600

Dd 1 MAGRLAGFLMLDGLASQSPAPACAGKMKVEEPMPTGLNNPFLPQASRLQPKREPSAVSG 600

Dy 61 PHLFRLSGKCFSLVESTYKKEFCEPHNVTQHQETPRMNAYSIGLGIHWEMELANNITG 120C
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 PLHLFRLAGKCFSLVESTYKKEFCEPHNVTQHQETPRMNAYSIGLGIHWEMELINNTFKG 120C

```
QY 121 MMKEDGDACRSRSRQSKVELACGKSNRLAHVSEPSVCYALTFETPLVCHPHALLVPEPTL 180C
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 MMKMTDGDGSGRSRSRQSKVELTCGKINRLAHVSEPSVCYALTFETPLVCHPHALLVPEPTL 180C
```

QY 161 PEALQQRQWQVEQDLADELTTPQGHKLRITLFEDAGYLKTPBENEPTQLGGSDSGFE 240
 ||||:|||||||:||||||| ||||| |||||
 Db 181 SEALQQRRLQVEQDLADELTTPQGYEKLRLVLFEDAGYLKVGETHPTQLAGSGSKGLGE 240

```
OY      241 TLNCRKAHNLSEIKRLKGLLTQHGIPYTPTETSNLEHLGHSTP-RAKSDEQLRSDP    299  
       ||:||||| ||::|| | ||||::||||::||| ::||| ||  
Db     241 TLDNCRKAAHELSGVTRLSSLLQGIGIPHOTPLTTHSQHQOOLPGALIAAENLRSDP   300
```

QY	300	GLRGSL	305
	:		
Db	301	GLRGNI	306

RESULT 13
US-09-986-552-9

```

: sequence 9, Application US/09386552
: Patent No. US20020150981A1
: GENERAL INFORMATION:
: APPLICANT: CANFIELD, William

```

FILE REFERENCE: 215089US77DIV
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL
CURRENT FILING DATE: 2001-11-09
CURRENT APPLICATION NUMBER: US/09/986,552

:
 : PRIOR APPLICATION NUMBER: 09/635,872
 :
 : PRIOR FILING DATE: 2000-08-10
 :
 : PRIOR APPLICATION NUMBER: 60/153,831
 :
 : PRIOR FILING DATE: 1999-03-14
 :

```

; NUMBER OF SEQ ID NOS: 52
;
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 9
;
; LENGTH: 307

```

```

; TYPE: PRT
; ORGANISM: Mus musculus
US-09-986-552-9

```

Query Match	78.9%;	Score 1288.5;	DB 9;	Length 307;
Best Local Similarity	78.4%;	Pred. No. 1.3e-121;		
Matches 240;	Conservative 25;	Mismatches 40;	Indels 1;	Gaps 1

DQ

1 MAAGLARIILLIGLSA GGPAPAGA AKKVV EEPNACGVNNPFL PQASRL QAKRDPSPISG 60
||| : ||| : ||| : ||| : ||| :
Db

1 MAGRIAGFIMLLGLS AQQGPAPA CAGKKVVEE PNTGLNPFL PQASRL QXPKEPSAVSG 60
|||| | : ||| : ||| : ||| : ||| :

QY 61 PVHLFRLSGKCSFVSTYKKEFCFPHNVTHQETPRMNAVSGILGIWHEWELANNFTG 120

Dp 61 PLHLFRLAGKCSFVSTYKKEFCFPHNVTHQETPRMNAVSGILGIWHEWELINNFTKG 120

```
QY      121 MMNFDGDACRSRQSKVELACGKNRLAHVSEPSTCVYALTFTETPLVCHPALLVPTL 180
        ||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dp      121 MMTDGDGSCHSRQSKVELTCGINRLAHVSEPSTCVYALTFETPLVCHPSLLVPTL 180
```

```

QY      16  PELLGQQMVQVQCDLADBLITPQGHKLLRTLPEDAGYLYKTPENEPQLGGPDSIGFE 240
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DQ      181  SEALQQRLLQVQVQCDLADBLITPQGYEKLLRVLPEDAGYLYKPGSTHTTQLAGSKXGLLE 240C
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY      241  TLENCRAKHELSKEIKRLKGLLTHQGIPIYRPTETSNLEHIGHETP-RAKSPQOLRGDP 239F
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DQ      241  TLQNCRAKHAELSGVEVQKRLTSLQGHGIPHTQPIETTHSHGLGQQLTGMALAEHLNSDP 300D
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY      300  GLRGSL 305
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DQ      301  GLRGNT 306
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

RESULT 14
US-10-023-888-12
Sequence 13 Application ref/10023888

Publication No. US20030119088A1
GENERAL INFORMATION:
APPLICANT: CANFIELD, William
TITLE OF INVENTION: CURVED CRANK PROSTHETIC MEMBER

```

; FILE REFERENCE: 203515U577
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21
; INVENTOR: CHEN, TAO, YONG

```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 307
;

```

ORGANISM: Mus musculus
US-10-023-888-12

Best Local Similarity 78.4%; Pred. No. 1.3e-121;
Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1

Db 1 MAGRLAGIMLTGLASOGGPAPACAGKMKVVEEPTFTGLNPFLLPQASRLQPREPSAVSG 60

61 P.H.I.F.R.L.A.G.K.C.F.S.V.E.S.T.Y.K.E.P.F.H.N.V.T.Q.H.E.Q.I.F.R.N.W.A.Y.S.G.I.L.G.I.W.H.E.W.E.I.N.N.T.F.K.G 120

Db

121 MNNITDGDSCHSRSGSKVELTGGKINRLHVSSESTCVVALLTFEPLVCHSHSLVYPTL 180

181 FEALQKQMDQVEQDLADSLITPGGKELKRLVLFEDAGYLKLFELNDFIQUNSGSFUSJFE 240
::: :::
181 SEALQGRDQVEQDLADSLITPGGKELKRLVLFEDAGYLKVPGETHPTQLAGGSKGGLGE 240

Dy 241 TLNCRKKAHLESLKEIKRKLGLTQHGIPYIRPIETSNLNLGHTEP-KAASPEQLGDP 299
||:|||||:::||:||||:||||:||||:||||:
Db 241 TLNCRKKAHLESLGEVGRILSLQOHGIPTHTQTFTTHSGHLGGQPLIGATAAEHLRSDP 300

QY	300	GLRGSL	305
	:		
Db	301	GLRGNI	306

RESULT 15
US-10-023-889-12
Sequence 12 Amplification ITS/10023889

Publication No. US20030124652A1
GENERAL INFORMATION:
APPLICANT: CANFIELD, William
TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMBINATION WITH A PROTEIN

```

; TITLE OF INVENTION: DEFICIENT CELLS
; FILE REFERENCE: 2035120577
; CURRENT APPLICATION NUMBER: US/10/023,889
; PENDING PILING DATE: 2001-12-31

```

```

; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin version 3.1

```

; SEQ ID NO 12
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-023-889-12

Query Match 78.9%; Score 1268.5; DB 14; Length 307;
Best Local Similarity 78.4%; Pred. No. 13e-12;
Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

QY	1	MAAGLARIILLGLSAGGPAPAGAAKMKVVEEPNAPGVNPPFLPQASRLQAKRDPSPVSG	60
DB	1	MAGRLAGFLMLGLASQGPAPACAGKMKVVEEPNTEGAINPPFLPQASRLQKREPSAVSG	60
QY	61	PVHLFRLSGKCFSLVESTYKYKFCFPHNVTOHEQTFRNNAYSGLIGIWHMEIANNTFTG	120
DB	61	PLHLFRLAGKCFSLVESTYKYKFCFPHNVTOHEQTFRNNAYSGLIGIWHMEIINNTEKG	120
QY	121	MMRPDGDACRSRSRQSKVELACGKSNRLAHVSEPTCYALTFFETPLVCHPHALLVYPTL	180
DB	121	MMMTDGDCHSRSRQSKVELTCGKINRLAHVSEPTCYALTFFETPLVCHPHALLVYPTL	180
QY	181	PEALQROWDQVEQDIADSEITPQGEKILRTLFEDAGYLKTPENEPFQLEGGPDSLGE	240
DB	181	SEALQORLDQVEQDIADSEITPQGEKILRVLFEDAGYLKVPGETHPTQLAGSGSKGLE	240
QY	241	TLENCRKAKHKLSEIKRLKGLLTQHGIPYTRPETSNTLHGHETP-RAKSPQLRGDP	299
DB	241	TIDNCRKAHAELSQEVORLTSILOQHGIPYTOPTEHTHSQHLGQQLPIGATAAHLRSDP	300
QY	300	GLRGSIL 305	
DB	301	GLRGNI 306	

Search completed: July 26, 2004, 11:25:26
Job time : 23.4465 secs

This Page Blank (uspio)

A:Residues: 1-1436 <STO>
A:Cross-references: GB:AE005176; PID:g12725306; PIDN:AAK06336.1; GSPDB:GN00146
A:Experimental source: strain IL403
C:Genetics:
A:Gene: yxca

Query Match 5.7%; Score 93.5; DB 2; Length 1436;
Best Local Similarity 27.3%; Pred. No. 21;
Matches 47; Conservative 21; Mismatches 55; Indels 49; Gaps 9;
QY 115 NNTFTGMRRDD--ACSRSRQSKVELACGKSNRLAHVSEPTCVYALTFETPL---- 167
DB 959 NKLIVYLVKIDEGSNMGAVIRLRSLKAAV---SERVRNITEASTVEHLYQETPAFTKE 1014
QY 168 VCHPVALVYPTLPALQKQWQVODLADLITPQGHKLTLPEDAGY--LKTPEEN 225
DB 1015 MAKKTLL---LP-----MLSTHQEGLLDTFAAAGNVAIVSLPESN 1053
QY 226 EPTQLEGPDLSGFETLNCRAKHELSKEIKELG-----LITQHG 267
DB 1054 --TSVNNG--LKFVNNDSCTPAITTIQGLIALQSGEYDLNNTSVMTQTG 1100

RESULT 9
T43152
hypothetical protein - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43152
R:Yoshioaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A:Reference number: Z17323; MUID:98162722; PMID:9501991
A:Accession: T43152
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-515 <YOS>
A:Cross-references: EMBL:D89245; NID:g1749697; PIDN:BAAL3906.1; PID:g1749698
A:Experimental source: strain PR745

Query Match 5.6%; Score 92; DB 2; Length 515;
Best Local Similarity 21.9%; Pred. No. 7;
Matches 25; Conservative 16; Mismatches 47; Indels 26; Gaps 3;
QY 87 HNVTHQEGTFR--NMAYSGLIGIWHMEIANNTFTGWM----- 122
DB 373 HTKMLHELYHNGWLYRAIKMETIKELIGTYVYEVVFPQDSILLGNFASQEGNVL 432
QY 123 -MRDGDAC-RSRSRQSKVELACGKSNRLAHVSEPTCVYALTFETPLVCHPHAL 174
DB 433 KYENQSGCWNGFHSALIVTEGCVENEIVSVLEAQCKEYLIMKSPACSPNQL 486

RESULT 10
136911
Involutrin L - douroucouli
C:Species: Aotus trivirgatus (douroucouli, night monkey, owl monkey)
C:Date: 16-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 22-Jun-1999
C:Accession: 136911
R:Tsang, H.; Green, H.
Mol. Biol. Evol. 6, 460-468, 1989
A:Title: The Involutrin gene of the owl monkey: origin of the early region.
A:Reference number: 136911; MUID:90014142; PMID:2507864
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-544 <RES>
A:Cross-references: GB:M25313; NID:g176557; PIDN:AAA5375.1; PID:g176558
C:Comment: During the terminal differentiation of keratinocytes, this protein from the c
linked envelope under the plasma membrane.
C:Superfamily: Involutrin
C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat
P:153-501/Region: 10-residue repeats (Q-E-G-Q-[PLV]-[KH]-[PL]-E-Q)

Query Match 5.6%; Score 92; DB 1; Length 544;
Best Local Similarity 23.4%; Pred. No. 7.5;
Matches 34; Conservative 25; Mismatches 48; Indels 38; Gaps 4;
QY 182 EALQKQWQVQDQDAE-----LITPQGHKLTLPED 215
DB 110 EKQAREKQQLQGLQEBEKKLLDQPDHLEAKSDEQLGTRKQDLPEPQEQQLKCLEQ 169
QY 216 AGYLTKEENEPYQLQEGGDSLGFETLNCRAKHELSKEIK-RKGLLTQHGIPYTRPT 274
DB 170 EGHLELPQEQEGO-----LKCLEQ-QEGHLELPQEQGLNHEQSQGLKHEQ 218
QY 275 ETSNLEHLEHTPRAPKSPQRLKDP 299
DB 219 QEGVYVHLEQEQEKSELPEQQRQGP 243

RESULT 11
C59436
KIAA1391 protein [imported] - human
C:Species: Homo sapiens (man)
C:Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 23-Sep-2002
C:Accession: C59436; D59436
R:Nagase, T.; Kikuno, R.; Ishikawa, K.I.; Hirose, M.; Ohara, O.
DNA Res. 7, 65-73, 2000
A:Title: Prediction of the coding sequences of unidentified human genes. XVI. The compl
A:Reference number: C59436
A:Accession: C59436
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1194 <NAG>
A:Cross-references: GB:BA92629; PID:g7243163; PIDN:BA92629.1
R:Ohara, O.; Nagase, T.; Kikuno, R.
submitted to Genbank, January 2000
A:Reference number: D59436
A:Accession: D59436
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1194 <OHA>
A:Cross-references: GB:BA92629; PID:g7243163; PIDN:BA92629.1

Query Match 5.6%; Score 92; DB 2; Length 1194;
Best Local Similarity 22.2%; Pred. No. 22;
Matches 63; Conservative 40; Mismatches 103; Indels 78; Gaps 13;
QY 14 LSAHGAPAPGA-----AKKYVEEPNMAFGVNNPFLQASRLQAR---DSEP-VSGPVH 63
DB 941 LSPPTSPSGSSVSSQDSAFQISEHSVFTPTETSSPIDCTFOAKRKEDLSPDSNASH 1000
QY 64 LFRLL---SGKCFSLVESYKYEF---CFHNVTQHEQTFRNMAVSGLIGIWHMEIANN 116
DB 1001 VSGMPGSSGQACSRPATYTKDVTMEHSGMSVTLHPST----- 1039
QY 117 FTFGMMRRDGA--CSRSRQSKVELACGKSNRLAHVSEPTCVYALTFETPLVCHPHAL 175
DB 1040 -----WLRNGVSLNKNWSIKKAKAPPEEEKIASPKXP-----LEPP--PHA-- 1080
QY 176 VFTPLPEALQKQWQVQDQDA--ADELTTPQGHKLTLPEDAGYLTKEENEPYQLQEGG 233
DB 1081 --SGVPEANSLQEQDPLPLRABGLSPVQAQROSSSPFDS-----ERHCSS 1127
QY 234 PSLGFETLNCRAKHELSKEIKELGILTLQHGIPYTRPTETS 277
DB 1128 PSLVESRLKLCMKSHHELEPQSQSSG-----SLPWRARASS 1166

RESULT 12
A33926
hypothetical protein 1 - chicken anemia virus
C:Species: chicken anemia virus, CAV
C:Date: 17-Jan-1992 #sequence_revision 17-Jan-1992 #text_change 08-Oct-1999
C:Accession: A33926


```

Db      976 APTKDIFLPYAEISRTYTKM-RNEASQILHTVE-TCHCFDKLSTNKLN---VESVTAD 1030
Qy      93  EQIFRNNAYSGLIGIMHEWEIANNTFTGMMWDDGACRSRSRQSKVELACGKNRLAHVS 152
Db      1031 ETT---DPASTLDLNNKESAGNESLEKOVFEDEVSSRQQLISTAGYLKCVQSN--LHIT 1084
Qy      153 EPSTCVYALT--ETPLVCHPHALIVPTLPEALOROWDOVEODLADLIT----- 201
Db      1085 VTSLVAAAYVWWSBPFLRN--ILP-LMASIKREQQLQIJAELAEALAYCVDR 1140
Qy      202 -PQGHKLLRT-----LFDAGYL---KTPENEPYOLEGGP 234
Db      1141 KPSPNDKLIKNICSLTCMDPSETPOASIISSMDIVDDMDFLSSRSNTGKQAKAVVLASGE 1200
Qy      235 DSL---GFETLENCRAHKELS-----KEIKRLKGLTQHGIP 269
Db      1201 DRSKYEGFTTRRGSELALKHLSLKFGGSIFDKLPKLMECTLEVLP 1246

```

Search completed: July 26, 2004, 11:12:26
 Job time : 9.22934 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:58 ; Search time 5.08008 Seconds
(without alignments)
3126.212 Million cell updates/sec

Title: US-10-657-280-3

Perfect score: 1634
Sequence: 1 MAAGLARRLLLLGASGPA.....TPRAKSPQRLGRDGLGSL 305

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155.5	9.5	667	1 OS9_HUMAN	Q1438 homo sapien
2	141	8.6	533	1 G19P_BOVIN	Q28034 bos taurus
3	140.5	8.6	527	1 G19P_HUMAN	P14314 homo sapien
4	113.5	7.1	377	1 K2L4_HUMAN	Q99706 homo sapien
5	95.5	5.8	216	1 VP2_CAVC1	Q99151 chicken ane
6	94.5	5.8	216	1 VP2_CAV82	P54093 chicken ane
7	92.5	5.7	216	1 VP2_CAV26	P54092 chicken ane
8	92	5.6	544	1 INV0_AOTTR	P24708 actus trivi
9	90	5.5	5262	1 ML22_HUMAN	O14686 homo sapien
10	89	5.4	2025	1 TTC3_HUMAN	P58804 homo sapien
11	88	5.4	593	1 STB2_CANFA	Q28288 canis faml
12	88	5.4	935	1 KINH_SYNRA	O43093 syncephalas
13	88	5.4	1328	1 YMD9_YEAST	Q03434 saccharomyc
14	87.5	5.4	364	1 Y9E8_PSEAE	P33642 pseudomonas
15	87.5	5.4	703	1 SYM_CHLTE	P59077 chlorobium
16	87	5.3	445	1 PLI3_MOUSE	Q88X76 mus musculu
17	87	5.3	469	1 PLI3_HUMAN	Q842N9 homo sapien
18	86.5	5.3	246	1 DEOC_THEAC	Q94kb7 thermoplasm
19	86.5	5.3	493	1 INVO_SAGOE	P24712 saguinus ce
20	85.5	5.2	560	1 INVO_PANPA	P14591 pan paniscu
21	85.5	5.2	797	1 STR3_HUMAN	Q15033 homo sapien
22	85.5	5.2	2128	1 SPGB_MOUSE	P15508 mus musculu
23	85	5.2	704	1 MYRB_MOUSE	P48972 mus musculu
24	84.5	5.2	4377	1 ANK3_HUMAN	Q13953 homo sapien
25	84	5.1	2871	1 DESP_HUMAN	P15924 homo sapien
26	83	5.1	1248	1 SM1B_MOUSE	Q920f6 mus musculu
27	82.5	5.0	286	1 PU91_SCICO	P22311 sciara copr
28	82.5	5.0	906	1 HELY_MYCTU	O10701 mycobacteri
29	81.5	5.0	322	1 PGJB_HUMAN	Q99645 homo sapien
30	81.5	5.0	383	1 PYRC_PYPAA	O84227 pyrobaculum
31	81.5	5.0	406	1 T233_HUMAN	Q96161 homo sapien
32	81.5	5.0	819	1 M4K2_HUMAN	Q18851 h mtogen-a
33	81	5.0	322	1 PGJB_MOUSE	P70186 mus musculu

34	81	5.0	579	1 MDIA_BUCBP	O69a97 buchnera ap
35	81	5.0	897	1 TBEC_HUMAN	O95759 homo sapien
36	81	5.0	1004	1 PTPX_RAT	Q63475 rattus norv
37	80.5	4.9	473	1 HYA2_RAT	Q924q3 rattus norv
38	80.5	4.9	835	1 INVO_PONPY	P14708 pongo pygma
39	80	4.9	279	1 ATPG_MYCPN	Q50330 mycoplasma
40	80	4.9	321	1 PG1B_BOVIN	P79119 bos taurus
41	80	4.9	410	1 VEE2_BPT1	P03122 bovine pap
42	80	4.9	770	1 GLOB_SYNY3	P52881 synchocyst
43	80	4.9	1328	1 YME4_YEAST	Q04711 saccharomyc
44	80	4.9	1328	1 YME5_YEAST	Q04214 saccharomyc
45	80	4.9	1328	1 YMU0_YEAST	Q04670 saccharomyc

ALIGNMENTS

RESULT 1
ID OS9_HUMAN STANDARD; PRT; 667 AA.
AC Q1438; 000579;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protein OS-9 precursor.
GN OS9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96198566; PubMed=8634085;
RA Su Y.A., Hutter C.M., Trent J.M., Meltzer P.S.;
RT "Complete sequence analysis of a gene (OS-9) ubiquitously expressed
RT in human tissues and amplified in sarcomas.";
RL Wcl. Carcinog. 15:270-275(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98158329; PubMed=9498564;
RA Kimura Y., Nakazawa M., Tsuchiya N., Asakawa S., Shimizu N.,
RA Yamada M.;
RT "Genomic organization of the OS-9 gene amplified in human sarcomas.";
RL J. Biochem. 122:1190-1195(1997).
RN [3]
RP SEQUENCE FROM N.A.; CHARACTERIZATION, AND ALTERNATIVE PRODUCTS.
RC TISSUE=Promyelocytic leukemia;
RX MEDLINE=98230694; PubMed=9562620;
RA Kimura Y., Nakazawa M., Yamada M.,
RT "Cloning and characterization of three isoforms of OS-9 cDNA and
RT expression of the OS-9 gene in various human tumor cell lines.";
RL J. Biochem. 123:876-882(1998).
RN [4]
RP SEQUENCE OF 194-667 FROM N.A.
RX MEDLINE=97336055; PubMed=9192850;
RA Elkahoun A.G., Krizman D.B., Wang Z., Hofmann T.A., Roe B.,
RA Meltzer P.S.;
RT "Transcript mapping in a 46-kb sequenced region at the core of 12q13.3
RT amplification in human cancers.";
RL Genomics 42:295-301(1997).
CC -!- FUNCTION: May play a role in cell viability.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=OS-9-1;
CC IsoId=Q1438-1; Sequence=Displayed;
CC Name=OS-9-2;
CC IsoId=Q1438-2; Sequence=VSP_004353;
CC Name=OS-9-3;
CC IsoId=Q1438-3; Sequence=VSP_004352; VSP_004353;
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed. Found as well in all
CC tumor cell lines analyzed, amplified in sarcomas. Highly expressed
CC in osteosarcoma OSA-CL and rhabdomyosarcoma RH30 cell lines. OS-9-
CC 2 is the major isoform detected in all cell types examined.

```

GN PRKCSH.
OS Bos taurus (Bovine) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
CX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Holstein; TISSUE=Corpus luteum;
RX MEDLINE=20149879; PubMed=10684806;
RA Brule S., Rahabi F., Faure R., Beckers J.-F.M.P., Silvestres D.W.,
RA Lussier J.G.,
RT "Vascular system-associated protein-60: a protein characterized from
RT bovine granulosa and luteal cells that is associated with
RT intracellular vesicles and related to human 80K-H and murine
RT beta-glucosidase II."
RL Biol. Reprod. 62:642-654(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U49178; AA92060.1; -
DR InterPro; IPR020248; EF-hand.
DR InterPro; IPR000886; ER target S.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR009011; Man_6_P_bind.
DR SMART; SM00182; LDla; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PROSITE; PS00018; EF_HAND; 1.
KW Phosphorylation.
KW DOMAIN
FT 313 334 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 533 AA; 60151 MW; 50795B677F8B1E01 CRC64;

Query Match 8.6%; Score 141; DB1; Length 533;
Best Local Similarity 30.7%; Pred. No. 0.00031;
Matches 35; Conservative 14; Mismatches 59; Indels 6; Gaps 4;

QY 67 LSGKCFELVSTYYEYECPPPHNTQHQGTFRMNAVSGILIGIHEWE-IANNPTGMMMRD 125
DB 416 LYSCEYLTLNNEYRYALCPFKLVSKR---KLGSPLSLGWSGMAQPDHDKPSAMKYEQ 472

QY 126 GDAC-FRSRSQSKVELAGCKSNRIAYVSEPTCYVALTFETPLVCHPALLVTP 178
DB 473 GTGCMQGNRSTTVRLICGKEVTVSTYESRCCEYLMLMTPAAC-PEPPEEYP 525

RESULT 3
G19P_HUMAN G19P_HUMAN STANDARD; PRT; 527 AA.
AC P14314;
DT 01-JAN-1990 (Ref. 13, Created)
DT 01-JAN-1990 (Ref. 13, last sequence update)
DT 15-JUN-1999 (Ref. 38, last annotation update)
DE Protein kinase C substrate, 80 kDa protein, heavy chain (PRKCSH)
DE (80K-H protein).
DE PRKCSH OR G19P1.
GN Homo sapiens (Human) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=900071553; PubMed=8793184;
RA Sakai K., Mesamichi H., Minoshima S., Kudoh J., Fukuyama R.,
RA Shimizu N.,
RT "Isolation of cDNAs encoding a substrate for protein kinase C:
RT nucleotide sequence and chromosomal mapping of the gene for a human
RT 80K protein."

```


RL Genomics 5:309-315 (1989).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97196775; PubMed=9043864;
 RA Ophoff R.A., Terwindt G.M., Verschuwe M.N., van Eijk R.,
 RA Mohrweiser H., Litz M., Hofker M.H., Haan J., Ferrari M.D.,
 RA Francis R.R.;
 RT "A 3-Mb region for the familial hemiplegic migraine locus on 19p13.1-
 RT p13.2: exclusion of PRKCSH as a candidate gene.";
 RL Eur. J. Hum. Genet. 4:321-328 (1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J03075; AAA52493.1; -
 DR EMBL: U50326; AAA98668.1; JOINED.
 DR EMBL: U50317; AAA98668.1; JOINED.
 DR EMBL: U50318; AAA98668.1; JOINED.
 DR EMBL: U50319; AAA98668.1; JOINED.
 DR EMBL: U50320; AAA98668.1; JOINED.
 DR EMBL: U50321; AAA98668.1; JOINED.
 DR EMBL: U50322; AAA98668.1; JOINED.
 DR EMBL: U50323; AAA98668.1; JOINED.
 DR EMBL: U50324; AAA98668.1; JOINED.
 DR EMBL: U50325; AAA98668.1; JOINED.
 DR PIR: A32469; A32469.1; JOINED.
 DR Genew: HGNC:9411; PRKCSH.
 DR MIM: 177060; -
 DR GO: GO:0005622; C:intracellular; NAS.
 DR GO: GO:0007243; P:protein kinase cascade; NAS.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000886; ER_target_5.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR009011; Man_6_P_R_bind.
 DR Pfam: PF00036; efhand; 2.
 DR SMART: SM00192; LDLA; 1.
 DR PROSITE: PS00014; ER_TARGET; 1.
 DR PROSITE: PS00018; EF_HAND; 1.
 KW Phosphorylation.
 FT DOMAIN 313 336 GLU-RICH (ACIDIC).
 FT MOD_RES 89 89 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 FT MOD_RES 382 382 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 FT MOD_RES 389 389 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 FT MOD_RES 433 433 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 SQ SEQUENCE 527 AA; 59296 MW; 2BB2F9AC14E3FC94 CRC64;
 Query Match 8.6%; Score 140.5; DB 1; Length 527;
 Best Local Similarity 30.5%; Pred. No. 0.00033;
 Matches 32; Conservative 14; Mismatches 54; Indels 5; Gaps 3;
 QY 67 LSGKCFSLVSTYKYPCEPHNVHTQETFMNAYSGLIGIHEW-BIANNFTGMWRD 125
 DB 410 LYSQCYELTNTYVRLCPFKVLSQK---KLGGPSLTSLGTGWSIGPDHDKFSAMKEQ 466
 QY 126 GDAC-RSRSRQSKVELLAKGSKNRLAHVSEPSVCYVALTFEPPLVC 169
 DB 467 GTCGQGFPRSTITVRLTCKGKMTVTSTSPSKCEYLMELTPAAC 511
 ID K214 HUMAN STANDARD; PRT: 377 AA.
 AC Q99706; O14621; O14622; O14623; O14624; O43534; P78400; P78401;
 AC Q99559; Q99560; Q99561; Q99562; Q90077;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 10-OCT-2003 (Rel. 42; Last annotation update)
 DE Killer cell immunoglobulin-like receptor 2D14 precursor (MHC class I

DE NK cell receptor KIR103AS) (Killer cell inhibitory receptor 103AS)
 DE (KIR-103AS) (G9P).
 GN KIR2DL4 OR KIR103AS.
 OS Homo sapiens (human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97102173; PubMed=8946682;
 RA Selvakumar A., Steffens U., Dupont B.;
 RT "NK cell receptor gene of the KIR family with two IG domains but
 RT highest homology to KIR receptors with three IG domains";
 RL Tissue Antigens 48:285-294 (1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 3; 4 AND 5).
 RX MEDLINE=97378880; PubMed=9234477;
 RA Selvakumar A., Steffens U., Palanisamy N., Chaganti R.S.K., Dupont B.;
 RT "Genomic organization and allelic polymorphism of the human killer
 RT cell inhibitory receptor gene KIR103.";
 RL Tissue Antigens 49:564-573 (1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 6).
 RC TISSUE=Lymphoid;
 RA Biasoni R.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. AND VARIANTS.
 RX MEDLINE=98090086; PubMed=9430221;
 RA Ueberberg M., Vallante N.M., Shum B.P., Shilling H.G.,
 RA Llenert-Weidenbach K., Corliss B., Tyran D., Lanier L.L., Parham P.;
 RT "Human diversity in killer cell inhibitory receptor genes.";
 RL Immunity 7:753-763 (1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Chan H.W., Salter R.D.;
 RT "Exon deletion contributes to structural diversity of 2D14 killer
 RT inhibitory receptors.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=97213129; PubMed=9059894;
 RA Selvakumar A.;
 RT "Polymorphism and domain variability of human killer cell inhibitory
 RT receptor.";
 RL Immunol. Rev. 155:183-196 (1997).
 CC -1- FUNCTION: RECEPTOR ON NATURAL KILLER (NK) CELLS FOR HLA-C ALLELES.
 CC INHIBITS THE ACTIVITY OF NK CELLS THUS PREVENTING CELL LYSIS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=6;
 CC Name=1;
 CC IsoId=Q99706-1; Sequence=displayed;
 CC Name=2; Synonyms=AST;
 CC IsoId=Q99706-2; Sequence=VSP_002609;
 CC Name=3; Synonyms=AS;
 CC IsoId=Q99706-3; Sequence=VSP_002610;
 CC Name=4; Synonyms=ASDI;
 CC IsoId=Q99706-4; Sequence=VSP_002609, VSP_002610;
 CC Name=5; Synonyms=ASD2;
 CC IsoId=Q99706-5; Sequence=VSP_002609, VSP_002610, VSP_002611;
 CC Name=6;
 CC IsoId=Q99706-6; Sequence=VSP_002608, VSP_002609, VSP_002610;
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC EMBL; U71199; AAB49756.1; -
DR EMBL; AF003123; AAB61926.1; -.
DR EMBL; AF003117; AAB61926.1; JOINED.
DR EMBL; AF003117; AAB61926.1; JOINED.
DR EMBL; AF003118; AAB61926.1; JOINED.
DR EMBL; AF003119; AAB61926.1; JOINED.
DR EMBL; AF003121; AAB61926.1; JOINED.
DR EMBL; AF003122; AAB61926.1; JOINED.
DR EMBL; AF003120; AAB61926.1; JOINED.
DR EMBL; AF002979; AAB71387.1; -.
DR EMBL; AF002980; AAB71387.1; -.
DR EMBL; AF002981; AAB71389.1; -.
DR EMBL; AF002982; AAB71390.1; -.
DR EMBL; X97229; CAA67842.1; -.
DR EMBL; X99479; CAA67842.1; -.
DR EMBL; X99480; CAA67843.1; -.
DR EMBL; AF034771; AAB95164.1; -.
DR EMBL; AF034772; AAB95165.1; -.
DR EMBL; AF034773; AAB95166.1; -.
DR EMBL; AF110035; AAD24763.1; -.
DR EMBL; AF110032; AAD24763.1; JOINED.
DR EMBL; AF110033; AAD24763.1; JOINED.
DR EMBL; AF110034; AAD24763.1; JOINED.
DR EMBL; U73384; AAC51146.1; -.
DR Gene; HGNC:6332; KIR2DL4.
DR MIM; 604945; -.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0004888; F:transmembrane receptor activity; TAS.
DR GO; GO:0006968; P:cellular defense response; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR007110; Ig-like.
DR Pfam; PF00047; Ig_2.
DR PROSITE; PS00835; IG-LIKE, FALSE NEG.
KW Receptor; Immunoglobulin domain; Glycoprotein; Signal Transmembrane;
KW Repeat; Multigene family; Alternative splicing; Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 377
FT DOMAIN 22 242
FT TRANSMEM 243 263
FT DOMAIN 264 377
FT DOMAIN 44 104
FT DOMAIN 139 202
FT DISULFID 51 97
FT CARBOHYD 146 195
FT CARBOHYD 141 141
FT CARBOHYD 175 175
FT VARSPLIC 27 121
FT VARSPLIC 219 235
FT VARSPLIC 236 270
FT VARSPLIC 271 288
FT VARSPLIC 53 53
FT VARSPLIC 87 87
FT VARSPLIC 138 138
FT VARSPLIC 209 209
FT VARSPLIC 271 271
FT VARSPLIC 371 371
FT CONFLICT 229 229
FT SEQUENCE 377 AA; 41527 MW; 7ECADF0FC7F4143B CRC64;

```

```

Query Match 7.1%; Score 115.5; DB 1; Length 377;
Best Local Similarity 21.3%; Pred. No. 0.027;
Matches 69; Conservative 43; Mismatches 117; Indels 95; Gaps 15;

QY 19 PAPAAGAKKVV--EEPNAFGV--NNPFLQARLQAKDPSVSGPV----- 62
DB 89 PAHAGTYRRCRGPHSPPTSMASNPVLVMVGLVEKPSLTARPPEPTVRTGENVTLLCSG 148
QY 63 -----HFRSLGKCFSL-----VESTYKEFCPPFNVQHQTFERNMAYSGILGIW 108
DB 149 QSSFDLYHLNR--EGEHEHLRLPAVPSINQTFQADP--PLGPAI--HGHTVR--CFGSPHSGP 203
QY 109 HEWEIANNFTFGKWMEDGACRSRQSVELACGSKNLAHVSEPTCVYLTETPLV 168
DB 204 YEWSDASDPL-----FVSVTGNPSSSWPSPPTBPS-----FKTGIA 236
QY 169 CHPH-----ALLVPTLPEALQRCQWQVEDDLADELITPGHEKLTLTLPEDAGYKLT 221
DB 239 RHHAHAYRVSVALIIFTLIPFLHRCCKKDAVMNQPAGHTVNR---EDSD----- 291
QY 222 PEENEPTLEGSDSIFFTLNCRAHKEISKIRLKGTLTGIGIPIYTR--TETSNLE 280
DB 292 -----EOPQOEVTYAQLDHCIFTRQKIRGSPQSRK-----RSTDTSVCI 331
QY 281 HIGHEPRAPKSPBQALRGDPRLGS 304
DB 332 ELPNAEPRLSPAHHEHSQLMGS 355

RESULT 5
ID VP2_CAVCI STANDARD; PRT; 216 AA.
AC Q99151; P54051;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Protein VP2.
OS Chicken anemia virus (German isolate Cuxhaven-1) (CAV), and
OS Chicken anemia virus (USA isolate CIA-1) (CAV).
OC Viruses; ssDNA viruses; Circoviridae; Gyrovirus.
OX NCBI_TaxId=73475; 73478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cuxhaven-1;
RX MEDLINE=91237831; PubMed=1851873;
RA Kraendburg O., Vos J.G., Jeurissen S.H.M., Hoebe R.C., Zanenema A.,
RA Koch G., van Ormondt H., van der Eb A.J.;
RT "Characterization of cloned chicken anemia virus DNA that contains
RT all elements for the infectious replication cycle.";
RL J. Virol. 65:3131-3139(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Cuxhaven-1;
RX MEDLINE=9226898; PubMed=1605740;
RA Meenan B.W., Todd D., Creelan J.L., Earle J.A.P., Hoey E.M.,
RA McNulty M.S.;
RT "Characterization of viral DNAs from cells infected with chicken
RT anaemia agent: sequence analysis of the cloned replicative form and
RT transfection capabilities of cloned genome fragments.";
RL Arch. Virol. 124:301-319(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CIA-1;
RX MEDLINE=97126092; PubMed=8971016;
RA Renshaw R.W., Soine C., Weinkle T., O'Connell P.H., Ohashi K.,
RA Watson S., Lucio B., Harrington S., Schat K.A.;
RT "A hypervariable region in VP1 of chicken infectious anemia virus
RT mediates rate of spread and cell tropism in tissue culture.";
RL J. Virol. 70:8872-8878(1996).
RN [4]
RP REVISION TO 24.

```

RA Renshaw R.W.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
 CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M55918; AAA91832.1; -
 DR EMBL; M81223; AAA42882.1; -
 DR EMBL; L14767; AAD09422.1; -
 DR PIR; A39926; A39926.
 DR PIR; A48343; A48343.
 DR InterPro; IPR006945; Circo_VP2.
 DR Pfam; PF04861; Circo_VP2; 1.
 DR CONFLICT 153 153 V -> A (IN REF. 1).
 FT CONFLICT 187 187 D -> N (IN REF. 2).
 SQ SEQUENCE 216 AA; 24138 MW; F82B69EF88BDE52 CRC64;
 Query Match 5.8%; Score 95.5; DB 1; Length 216;
 Best Local Similarity 19.6%; Pred. No. 0.65;
 Matches 54; Conservative 44; Mismatches 74; Indels 103; Gaps 14;
 QY 17 GGPAPAGAAKMKVVEEPNAPFGVNNPFLPQASRLQAKRDPSPVSGPVHLFRLSGKCFSLYE 76
 DB 21 GQPGPSGAAGQGVISN-----ERSR-----RSTRITNGVQ 52
 QY 77 STYKYECPFHNTQHEQTFRNAYSGILGIW-----HEWEIAN-NFTFGMMRDGDAC 129
 DB 53 ATKKFTAVGNPSIQRPDPWYRWN-YNHSIAVWLRECSRHAKICNGQFRKXWFOECAGL 111
 QY 130 RRSRSQSKVELACGSKNRLA-----HYSEPTCYALTFTFPLVCHPHALLVYPT 179
 DB 112 EDRSTQASLEBALRLPRLVQGRKAKRLDYHSQPT-----PNKKKYKT 156
 QY 180 LPEALQROWQVQDADLDELITPQGEKILRTLFEDAGYLKTPENEPPTOLEGGPDSLGF 239
 DB 157 V-----RW-----KDELAD-----READF--TPSER--DGGTSSDF 164
 QY 240 ETLNCRRAKHELSKEIKLGLTQ-HGIPYTRP 273
 DB 185 D-----EDINFDIGDSGIVDELLGRPFTTP 210
 RESULT 6
 VP2_CAV82
 ID VP2_CAV82 STANDARD; PRT; 216 AA.
 AC P54093;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Protein VP2.
 OS Chicken anemia virus (Japanese isolate 82-2) (CAV).
 CC Viruses; ssDNA viruses; Circoviridae; Gyrovirus.
 CC NCBI_TaxID=73476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95297149; PubMed=7778281;
 RA Kato A., Fujino M., Nakamura T., Ishihara A., Otake Y.;
 RT "Gene organization of chicken anemia virus.";
 RL Virology 209:480-488(1995).
 CC -!- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
 CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D31965; BAA06732.1; -
 DR InterPro; IPR006945; Circo_VP2.
 DR Pfam; PF04861; Circo_VP2; 1.
 SQ SEQUENCE 216 AA; 24138 MW; E23169F5925DD52 CRC64;
 Query Match 5.8%; Score 94.5; DB 1; Length 216;
 Best Local Similarity 19.6%; Pred. No. 0.78;
 Matches 54; Conservative 44; Mismatches 74; Indels 103; Gaps 14;
 QY 17 GGPAPAGAAKMKVVEEPNAPFGVNNPFLPQASRLQAKRDPSPVSGPVHLFRLSGKCFSLYE 76
 DB 21 GQPGPSGAAGQGVISN-----ERSR-----RSTRITNGVQ 52
 QY 77 STYKYECPFHNTQHEQTFRNAYSGILGIW-----HEWEIAN-NFTFGMMRDGDAC 129
 DB 53 ATKKFTAVGNPSIQRPDPWYRWN-YNHSIAVWLRECSRHAKICNGQFRKXWFOECAGL 111
 QY 130 RRSRSQSKVELACGSKNRLA-----HYSEPTCYALTFTFPLVCHPHALLVYPT 179
 DB 112 EDRSTQASLEBALRLPRLVQGRKAKRLDYHSQPT-----PNKKKYKT 156
 QY 180 LPEALQROWQVQDADLDELITPQGEKILRTLFEDAGYLKTPENEPPTOLEGGPDSLGF 239
 DB 157 V-----RW-----KDELAD-----READF--TPSER--DGGTSSDF 164
 QY 240 ETLNCRRAKHELSKEIKLGLTQ-HGIPYTRP 273
 DB 185 D-----EDINFDIGDSGIVDELLGRPFTTP 210
 RESULT 7
 VP2_CAV26
 ID VP2_CAV26 STANDARD; PRT; 216 AA.
 AC P54092;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Protein VP2.
 OS Chicken anemia virus (USA isolate 26p4) (CAV).
 CC Viruses; ssDNA viruses; Circoviridae; Gyrovirus.
 CC NCBI_TaxID=73477;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341490; PubMed=1908516;
 RA Claessens J.A.J., Schrier C.C., Mockett A.P.A., Jagt E.H.J.M.,
 RA Sondermeijer P.J.A.;
 RT "Molecular cloning and sequence analysis of the genome of chicken
 RT anaemia agent";
 RL J. Gen. Virol. 72:2003-2006(1991).
 CC -!- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
 CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10068; BAA00959.1; -
 DR InterPro; IPR006945; Circo_VP2.
 DR Pfam; PF04861; Circo_VP2; 1.
 SQ SEQUENCE 216 AA; 24066 MW; CD4B69EF88BDDC72 CRC64;
 Query Match 5.7%; Score 92.5; DB 1; Length 216;
 Best Local Similarity 19.6%; Pred. No. 1.2;
 Matches 54; Conservative 43; Mismatches 75; Indels 103; Gaps 14;

```

QY 17 GGGPAGAAKMKVVEBPNAFGVNNFLPQASRLAKRDPSPVSGVHFLRSGKCFSLVE 76
DB 21 GPGSGAAAGGVIN-----ERSPR-----RSTTINGVQ 52
QY 77 STYKTEFCFENHNTQEQTFKMNAYSGILGIM-----HEWEIAN-NTFTGMNRDGDAC 129
DB 53 ATNKFPAVGNPSLQDPPMYRW-NYHSLAVLRECSRSKAKI CNGCGFRGHMFQECAGL 111
QY 130 RRSRSOSKVELACGSNRLA-----HYSESTCVALLTETPLVCHPHALLVYPT 179
DB 112 EDRSTQASLEELILPLFVQGRARAKLDYHISQPT-----PNEKKVYKT 156
QY 180 LPEALQROMDQVEODLADLITPQGHKTLRLTFEDAGYLTPEENPTQLGGPDSIGF 239
DB 157 V-----RW-----QDELAD-----READF--TPSBE-----DGGTTSDF 184
QY 240 EYLENCRKAHKLSEIKRLKGLTLQ-HGIPYTP 273
DB 185 D-----GDINPDIGDSGIVELLGRFPTT 210

RESULT 8
ID INVO_AOTTR STANDARD; PRT; 544 AA.
AC P24708;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Involutrin.
GN IVL.
OS Aotus trivirgatus (Night monkey) (Douroucouli).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX NCBI_Taxid=9505;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90014142; PubMed=2507864;
RA Tseng H., Green H.;
RT "The Involutrin gene of the owl monkey: origin of the early region.";
RL Mol. Biol. Evol. 6:460-468(1989).
CC -1- FUNCTION: Part of the insoluble cornified cell envelope (CE) of stratified squamous epithelia.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding of the cornified envelope.
CC -1- TISSUE SPECIFICITY: Keratinocytes of epidermis and other stratified squamous epithelia.
CC -1- PTM: Substrate of transglutaminase. Specific glutamines or lysines are cross-linked to keratins, desmoplakin and to inter involutrin molecules.
CC -1- SIMILARITY: Belongs to the involutrin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M25313; AAA5375.1; -
CC DR PIR; I36911; I36911.
CC DR InterPro; IPR002360; Involutrin.
CC DR InterPro; IPR000354; Involutrin_rpt.
CC DR Pfam; PF00904; Involutrin_39.
CC DR PROSITE; PS00795; INVOLUCRIN; 1.
CC KM Keratinization; Repeat.
CC SQ SEQUENCE 544 AA; 63927 MW; 2A02ABR5E1499F9D CRC64;

Query Match 5.6%; Score 92; DB 1; Length 544;
Best Local Similarity 23.4%; Pred. No. 4.2;
Matches 34; Conservative 25; Mismatches 48; Indels 38; Gaps 4;
QY 182 EALQROMDQVEODLADLITPQGHKTLRLTFEDAGYLTPEENPTQLGGPDSIGF 215
-----LITPQGHKTLRLTFED 215

```

```

DB 110 EKAQRKQQLQGLQLEEKELLDQQRDHLSKSDQELGTRKQQLLFFPQEQQLKCLBQ 169
QY 216 AGYLTPEENPTQLBGGPDSIGFTELNCRKAHKLSEIKR-RKGLTLTGITPTPT 274
DB 170 EGHLELPEQEQEQ-----LKCLBQ-QEGHLEPQEQEQGLKLEQEGQLKLEQ 218
QY 275 ETSNLEHLGHETPRAKSPPEQLGDP 299
DB 219 QEGQVHLEQEQEQSELPEQQRGP 243

RESULT 9
ID ML12_HUMAN STANDARD; PRT; 5262 AA.
AC 014686; 014687;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related protein).
GN ML12 OR ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with RT strong homology to ALL-1 involved in acute leukemia and to Drosophila RT trithorax.";
RL Oncogene 15:549-560(1997).
RN [2]
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-N., Kang M.-J., Jung D.-J.,
RA Kvak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Aotora D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-D., Lee Y.C.,
RA Lee J.W.;
RT "activating signal cointegrator 2 belongs to a novel steady-state RT complex that contains a subset of trithorax group proteins.";
RL Mol. Cell. Biol. 23:140-149(2003).
CC -1- FUNCTION: May be involved in transcriptional regulation.
CC -1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/ CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins CC ML12 and ML13, and ASH2/ASCL2.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=O14686-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O14686-2; Sequence=VSP_008563, VSP_008559;
CC Name=3;
CC IsoId=O14686-3; Sequence=VSP_008560;
CC -1- TISSUE SPECIFICITY: Expressed in most adult tissues, including a CC variety of hematopoietic cells, with the exception of the liver.
CC -1- MISCELLANEOUS: This gene mapped to a chromosomal region involved CC in duplications and translocations associated with cancer.
CC -1- SIMILARITY: Belongs to the transcription factor trithorax family.
CC -1- SIMILARITY: Contains 5 PHD-type zinc fingers.
CC -1- SIMILARITY: Contains 1 post-SET domain.
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
CC -1- SIMILARITY: Contains 1 SET domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration - CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

Query Match	5.5%; Score 90; DB 1; Length 5262;
Matches	72; Conservativity 18.9%; Pred. No. 1.2e+02; Mismatches 145; Indels 128; Gaps 15;
15 SAGGPAPAGAAKMK-----	YVEENARAGVNNPFLPQASRLQAKRDPVS--- 59
70 NCGEPALHGGRELRFELPDMWPCPVVSGSGGPRNEAVLP-----	SEDSLQIGPPE 122
60 --GVHLFRLSGKCFSLVESTKYKFCFPHNVTOHEQTFMNAYSGILGIMH-----	WE 112
123 GLTPAHGHPGSGCMA-----	HHWCAAMSA-----GWGQGPGQLCG 159
113 IANNFTFGMMRRDDA-----	CRGRS-----RQSVETLACGSKNR 147
160 VDKAIFSGISGRCHCHTRLGASITCRSPGCRPLVHPCFCAIASGFSMKTLLDLPCHHSZ	219
148 LAHVSSESTCYVA-----	LTETPLVCHPHALLVYPTLPEALQWQMDVQDLDEL 199
220 GAALIEARCAVCEGPELCLDFECTCGHNYHGACIDTALTARKRAGMCCPECKVQAC	279
200 ITPGCHKRLRTLPEDAGY-----	LKTEENHPT-----QLEGGRSLG 238
280 RKPNDBSKMLVCELCGKGYHTFLCKLPMEELPAHSWKCKACRYCRACGASAEILNPVSEW	339
239 FETLENCRAKHAKELSKIKRLKGLLTOH-----	GLPYRTETETSLNHL----- 282
340 FENSYLCHRGCK--AOGGQITRSVAEQHTVCSRFSPPEGDPPTDEPDALVYACQGP	397
283 -GHET---PRAKSPPELRGDP 299	
398 GGHVTSMQPKPEPFLQCEAKP 418	

DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tetratricopeptide repeat protein 3 (TPR repeat protein D).
 GN TTPC3 OR TTPD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxID=9606;
 RX MEDLINE=96281123; PubMed=8724848;
 RA Chiba M., Ootsuyama A., Suzuki E., Ichikawa H., Seki N.,
 RA Nagase T., Nomura N., Ohki M.,
 RT "Identification of a novel human gene containing the
 RT tetratricopeptide repeat domain from the Down syndrome region of
 RT chromosome 21.";
 RL DNA Res. 3:9-16(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97103476; PubMed=8947847;
 RA Teubahara F., Hattori M., Muraki T., Sakaki Y.;
 RT "Identification and cloning of a novel cDNA belonging to
 RT tetratricopeptide repeat gene family from Down syndrome-critical
 RT region 21q22.2.";
 RL J. Biochem. 120:820-827(1996).
 RN [3]
 RP SEQUENCE OF 121-616 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=98163741; PubMed=9503011;
 RA Dahmane N., Alt Ghezala G., Gossel P., Chamoun Z.,
 RA Dufresne-Zacharia M.C., Lopes C., Rabatel N., Gassanova-Maugenre S.,
 RA Chetouh Z., Abramowski V., Fayet E., Yaspo M.L., Korn B.,
 RA Blouin J.L., Leheirch H., Pousetka A., Antonarakis S.E., Sinet P.M.,
 RA Creau N., Delabar J.M.;
 RT "Transcriptional map of the 2.5-Mb CBR-ERG region of chromosome 21
 RT involved in Down syndrome.";
 RL Genomics 48:12-23(1998).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=TRPD1;
 CC IsoId=PS3804-1; Sequence=Displayed;
 CC Name=TRPD1;
 CC IsoId=PS3804-2; Sequence=VSP_006554;
 CC Name=TRPD1;
 CC IsoId=PS3804-3; Sequence=VSP_006555;
 CC -1- TISSUE SPECIFICITY: Found in all tissues examined.
 CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D83077; BAA11769.1; -
 DR EMBL; D84294; BAA12301.1; -
 DR EMBL; D84295; BAA12302.1; -
 DR EMBL; D84296; BAA12303.1; -
 DR EMBL; A7001866; CAA05057.1; -
 DR PIR; JCS020; JCS020
 DR Gene; HGNC:12593; TTC3.
 DR WIM; 602259; -
 DR GO; GO:0009405; P:pathogenesis; TAS.
 DR InterPro; IPR008940; Premyl_trans.
 DR InterPro; IPR001440; TPR.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00515; TPR; 3.

DR Pfam; PF00097; zf-C3HC4; 1.
 DR SMART; SM00184; RING; 1.
 DR SMART; SM00028; TPR; 3.
 DR PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
 DR PROSITE; PS00089; ZF_RING_2; 1.
 KW Repeat; TPR repeat; Zinc-finger; Alternative splicing; Polymorphism.
 FT REPEAT 231 264
 FT REPEAT 266 298
 FT REPEAT 536 572
 FT REPEAT 576 609
 FT ZN_RING 1957 1997
 FT DOMAIN 453 456
 FT DOMAIN 1020 1029
 FT DOMAIN 1899 1902
 FT DOMAIN 1018 1029
 FT DOMAIN 1172 1185
 FT DOMAIN 1563 1579
 FT VARSPPLIC 1 233
 FT VARSPPLIC 1 310
 FT VARSPPLIC 1 310
 FT VARIANT 1117 1117
 FT CONFLICT 121 121
 FT CONFLICT 139 139
 FT CONFLICT 232 232
 FT CONFLICT 276 276
 SQ SEQUENCE 2025 AA; 229889 MW; 1B4BCAA3684B6253 CRC64;
 Query Match 5.4%; Score 89; DB 1; Length 2025;
 Best local Similarity 21.5%; Pred. No. 41;
 Matches 65; Conservative 46; Mismatches 123; Indels 68; Gaps 14;
 QY 26 KKKVEEPNAPGVNPNFLQASRLQ--AKRDPG--PVGG-VHLFRLSKGCEFLVST 78
 DB 1325 RPDITQPPAYINIVAGLPQYISYIYPLASLSPYQLPSPVSVPSFVANDADKNAAY 1384
 QY 79 YKYEPCFFINVOQHEQTFPMNAVSGILGFWHEMEIANNFTMMNRDGDACRSRROSKV 138
 DB 1385 FEGHLMNAENVAGHQIASETOILEGSLG---SVKSHOST-----GDHVLVSESNRN 1434
 QY 139 ELAAGKSNRIAHYSEPSCTVVALTETPLVCPHALLVYPTLPALQROMQVODDLADE 198
 DB 1435 DEHCGSNMKCEVTFPEST-----SAVTNIPHQMV-----AIQVSWNIHQEVNTE 1480
 QY 199 LIITP-----CG-----HEKL---LRTLFEDAGVILKTP-----EENB- 226
 DB 1481 PYNPFEEQGEISRIKEHQVLODQLOEVYENVEQIKLKGLETRDLDEKLRHLLENKI 1540
 QY 227 -PYLEGGPDLSGFTLENCRAKELSKIRLKLGL---TQHGIPYTRPTETSNLEHL 282
 DB 1541 SKTELWMPQLDLERE-IKKMQGKKKIQERLSLKKIKKIVSNASPMYQKNDGKEKEHE 1599
 QY 283 GH 284
 DB 1600 LH 1601
 RESULT 11
 STB2_CANFA
 ID STB2_CANFA STANDARD; PRT; 593 AA.
 AC Q28288;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Syntaxin binding protein 2 (Unc-18 homolog 2) (Unc-18B) (Unc-18-2).
 GN STXB2
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 NC NCBI_TaxID=9615;
 RX [1]
 RP SEQUENCE FROM N.A.

CC TISSUE=Kidney;
 CC MEDLINE=96370811; PubMed=8774707;
 RA Riento K., Jaenit J., Jansson S., Hielm S., Lehtonen E., Enholm C.,
 RA Keränen S., Olkkinen V.M.,
 RT "A secret-related vesicle-transport protein that is expressed
 RT predominantly in epithelial cells."
 RL Eur. J. Biochem. 239:638-646(1996).
 CC -!- FUNCTION: Involved in the protein trafficking from the Golgi
 CC apparatus to the plasma membrane (By similarity).
 CC -!- SUBUNIT: Binds syntaxins 1A, 2, 3 but not 4 (By similarity).
 CC -!- SIMILARITY: Belongs to the STXB/UNC-18/Sec1 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L41609; AAC00031.1; -!
 DR InterPro; IPR001619; Sec1-like.
 DR Pfam; PF00095; Sec1; 1.
 KM Protein transport.
 SQ SEQUENCE 593 AA; 66499 MW; BC42120B638FC737 CRC64;
 Query Match 5.4%; Score 88; DB 1; Length 593;
 Best Local Similarity 21.4%; Pred. No. 10;
 Matches 61; Conservative 33; Mismatches 99; Indels 92; Gaps 13;
 QY 18 GAPAPGAKMKVKEPNFAGVNNPRL---PQASRLQ---ARRDSPPVGGPRLRLSKGC 71
 DB 134 GPEVTAQLANAVLALNFKADNPSLGSEPKTRSLQLLIVRGADPVSPLLEHLPQAMA 253
 QY 72 FSLV---ESTYKYE-----
 DB 254 YDLNLIEDQTYREYETTGSEAREKAVLDEDDDLVELRHMHADVSKVTELLKTCES 313
 QY 87 HNVTHGEQTFRNNAVSLIGIWEHETIANNTFTGMKWDGACSRSSQSVIELACGSKN 146
 DB 314 KRLTIDDKANIK--DLSHILKMPQYQKELNKYS--THLHADDCKMHFKGS--VEKLCGVEQ 369
 QY 147 RLAVHSEPTCVYALTFFPLVCHPFLVYPTLPEALOROMDOY-----E 192
 DB 370 DLAMGSDT-----EGEKIKDMKILV--PVLDAVAPYDKIRVILLYILLRNGVSE 419
 QY 193 QDLADELT---TPQGEKILRTLPFDAGYLTKEPNEPTLEGGP 234
 DB 420 EMLA-KLIOHNAVQAHSSILRLIEQLGCTVNP-----GGP 454
 RESULT 12
 KINH_SYNRA
 ID KINH_SYNRA STANDARD; PRT; 935 AA.
 AC 043053;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Kinesin heavy chain (Synkin).
 OS Syncephalastrum racemosum.
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales;
 OC Syncephalastraceae; Syncephalastrum.
 OX NCBI_TaxID=13706;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98273660; PubMed=9613604;
 RA Grunert M., Piecor S., Iotsepeich F., Schliwa M.;
 RT "Cloning and functional expression of a 'fast' fungal kinesin."
 RL FEBS Lett. 427:79-84(1998).
 CC -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
 CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. ITS MOTOR
 CC ACTIVITY IS DIRECTED TOWARD THE MICROTUBULE'S PLUS END. THE SPEED
 CC OF THIS MOTOR IS 4-5 TIMES FASTER THAN ITS ANIMAL COUNTERPARTS.

CC -!- DOMAIN: Composed of three structural domains: a large globular N-
 CC terminal domain which is responsible for the motor activity of
 CC kinesin (it hydrolyzes ATP and binds microtubule), a central
 CC alpha-helical coiled coil domain that mediates the heavy chain
 CC dimerization; and a small globular C-terminal domain which
 CC interacts with other proteins (such as the kinesin light chains),
 CC vesicles and membranous organelles.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family. Kinesin
 CC subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AJ25894; CA12647.1; -!
 DR PIR; T51930; T51930.
 DR HSP; P3176; 1BQ2.
 DR InterPro; IPR007420; DUF465.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF04325; DUF465; 1.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PSS0667; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil.
 FT DOMAIN 1 341 KINESIN-MOTOR (BY SIMILARITY).
 FT DOMAIN 342 887 COILED COIL (POTENTIAL).
 FT DOMAIN 888 934 GLOBULAR.
 FT NP BIND 87 94 ATP (POTENTIAL).
 FT NP BIND 237 244 ATP (POTENTIAL).
 SQ SEQUENCE 935 AA; 105040 MW; CFBDP33B3827E28 CRC64;
 Query Match 5.4%; Score 88; DB 1; Length 935;
 Best Local Similarity 25.0%; Pred. No. 18;
 Matches 27; Conservative 27; Mismatches 34; Indels 20; Gaps 3;
 QY 178 PLPEALOROMDVEDDLADELTLPQ---GHEKILRTLPFDAGYLTKEPNEPTLEGGP 224
 DB 419 PLPEKDERREBFIRKEMLDQISSEKTELTNREKLLSREMGVYKEPQSVYKENQM 478
 QY 225 NEPTLEGGPDLSGFTLENC-----RKARELSKEIKRLKGLTQ 265
 DB 479 TSELSELRLQLQKVSYESKENAIVDSLKEANODLMAEIEIKNLISE 526
 RESULT 13
 YMD9_YEAST
 ID YMD9_YEAST STANDARD; PRT; 1328 AA.
 AC Q03434;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transposon Ty1 protein B.
 GN TY1B OR YML035W OR YMO054.04.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288c / AB972;
 RX MEDLINE=97313268; PubMed=9169872;
 RA Bowman S., Churche C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagsels K., Iye G., Moul S., Odell C., Pearson D., Rajandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrett B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT XIII."

RL Nature 387:90-93(1997).
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, Z48430; CAA88330.1; -
 DR PIR: S52481; S52481.
 DR SGD: S0004503; YML039W.
 DR InterPro: IPR001969; Asparticase_AS.
 DR InterPro: IPR009007; Pept_A_acid.
 DR InterPro: IPR001584; Rev.
 DR Pfam: PF00665; rve; 1.
 DR PROSITE: PS00141; ASP_PROTEASE; 1.
 KW Transposable element; Hypothetical protein; Hydrolase;
 KW Aspartyl protease; ATP-binding
 FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
 FT NP_BIND 1204 1211 ATP (POTENTIAL).
 SQ SEQUENCE 1328 AA; 151036 MW; C7D14E1A4A675E93 CRC64;
 Query Match 5.4%; Score 88; DB 1; Length 1328;
 Best Local Similarity 22.9%; Pred. No. 29;
 Matches 77; Conservative 34; Mismatches 85; Indels 140; Gaps 22;
 QY 55 PEPVSGP---VHLRLSGKCSLVESTYKTEFCFENVTOG--EQTFRNAYSGILGIM 108
 DB 146 PENVISVPTNNVH-----TSESTRKYVPPIHMLHANAQTIRYS----- 186
 QY 109 HEMELIANFTGMWRDGDACRSRQSKVELAC-----GKSNRLAHV----- 151
 DB 187 ---LKNNTIT--YNESEDVDRS---SAIDQCDCDIGSTKRHKSGRLKQNSYE 236
 QY 152 -----SEPTCYVALTFET-----PLVCHPH---ALLVYPTLPE 182
 DB 237 PQYLTHTDIFGPVHNLPSKSPSYFI-SPTDETFKRWYPL--HNRREDSTLDVFTTILA 293
 QY 183 ALQQRQWDQ---VEQDLADEL-----ITP-----QGH---EKLRLTLE 214
 DB 294 FIKNGPQASVLVIQMDRGSEYNNRLHFKLEKNGITPCYTTAASRAICVAERLNLTD 353
 QY 215 DAGYLTKEPENEPTOLE--GGPDSLGFTLENCRAKHEL-----SKEIKR---LKGLLT 264
 DB 354 DC-----RTQCGSGGLPNNHMFSAIEPSTIVRNSLASPKSKGAPQAGIAGUDI 403
 QY 265 QHGIPTRPTETSNLEHGHETPRAKSPBQLRGDPG 300
 DB 404 STLLPFGQPIVND-----HNPNSKI--HPRGIFG 431
 RESULT 14
 Y9B8_PSEAR STANDARD; PRT; 364 AA.
 ID Y9B8_PSEAR Q51527; Q9HVN1;
 AC P33642; Q51527; Q9HVN1;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable D-amino acid oxidase PA4548.
 GN PA4548.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RX MEDLINE=96272255; PubMed=8682785;
 RA Alm R.A.; Mattick J.S.;
 RT "Identification of two genes with prepilin-like leader sequences

RT involved in type 4 fimbrial biogenesis in Pseudomonas aeruginosa.";
 RL J. Bacteriol. 178:3809-3817(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Strover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Matman S., Yuan X.,
 RA Brody L.L., Collier S.N., Folger K.R., Kas A., Lavig K., Lim R.W.,
 RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 RN [3]
 RP SEQUENCE OF 193-364 FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=9325810; PubMed=8097014;
 RA Hobbs M., Collier E.S.R., Free P.D., Livingston S.P., Mattick J.S.;
 RT "Pils and PilR, a two-component transcriptional regulatory system
 RT controlling expression of type 4 fimbriae in Pseudomonas
 RT aeruginosa.";
 RL Mol. Microbiol. 7:669-682(1993).
 CC -1- COFACTOR: FAD (Potential).
 CC -1- SIMILARITY: Belongs to the dadA family of oxidoreductases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, I48934; AAB39269.1; -
 DR EMBL: AE004688; AAG07936.1; -
 DR EMBL: Z12154; CAI78142.1; -
 DR PIR: B83078; B83078.
 DR InterPro: IPR006076; Fad_oxred.
 DR InterPro: IPR000205; NAD_BS.
 DR InterPro: IPR000594; Thif_domain.
 DR Pfam: PF01266; DAO_1.
 KW Hypothetical protein; Oxidoreductase; Flavoprotein; FAD;
 KW Complete proteome.
 FT NP_BIND 5 19 FAD (ADP PART) (POTENTIAL).
 FT CONFLICT 20 20 E -> K (IN REF. 1).
 FT CONFLICT 23 23 L -> I (IN REF. 1).
 FT CONFLICT 57 57 S -> N (IN REF. 1).
 FT CONFLICT 109 109 R -> P (IN REF. 1).
 FT CONFLICT 193 245 IRGDKYLAAGAMSGELIKPGLIEIPVYVKGOMITLYKCA
 FT DFLPRVTLAKGR -> DPKRKAGAGGRLEKRYVEAWPMT
 FT ARGTERSDPQLQVKGGFPAAGAGQGG (IN REF. 3).
 FT ASA -> VSV (IN REF. 1).
 FT CONFLICT 280 282
 SQ SEQUENCE 364 AA; 39445 MW; 86C82DC62AA7A811 CRC64;
 Query Match 5.4%; Score 87.5; DB 1; Length 364;
 Best Local Similarity 20.4%; Pred. No. 6;
 Matches 56; Conservative 28; Mismatches 77; Indels 113; Gaps 11;
 QY 23 GAAMKVVSEPNFAFGVNNPFLPQASRLQAKRDPSPVGFVHLRLSGKCSLVESTYKXE 82
 DB 131 GAGFGRAVYVMSGVANVRNRLAKSLASIQQ----- 161
 QY 83 FCFPHAVTQHEQFFRNAYSGILGTHHEEINANTFTGMMDGD-----ACRSRQS 136
 DB 162 ---FANLEDEHQT-----EVGG-----WLKDGRVGVATSRSEIKRD 196
 QY 137 KVELIACGK-SNRLAHVSEPTCYVALTFETPCHPALVYPTPEALQQRQWDVEDPL 195
 DB 197 KYLLAAGAMSGEL-----LKPLGLELPVYVYVKGOMITLYKCAADFLPR----- 238
 QY 196 ADELITPQGH-----EKLRLTLEDAAGYLTKEPENEPTQLEGPSLSGFTLENCRA 247


```

DR 239 ---MWLAKRVAIPRRDGHILGSLTHSGFDPKXTD-----EAGESURA 280
QY 248 AHKEISEKEKRLK-----GLL--TQHGIPYTRP 273
DB 281 SAAEILPELADMQPVAHWAGLRPGSPGPIYIGP 314

RESULT 15
SYM_CHLITE
ID SYM_CHLITE STANDARD: PRT; 703 AA.
AC P59077;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
DE (MeRS).
GN METG OR METS OR CT0969.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_Taxid=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eissen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Debey R., Gwin M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Unayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nietman W.C., Feldblyum T.V., Hansen C.L., Craven W.B., Radune D.,
RA Vamathevan J., Khouri H., Bryant D.A., Fraser C.M.;
RA "The complete genome sequence of Chlorobium tepidum TLS, a
RA photosynthetic, anaerobic, green-sulfur bacterium.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
RL
RT FUNCTION: Is required not only for elongation of protein synthesis
RT but also for the initiation of all mRNA translation through
RT initiator tRNA (Met) aminoacylation (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
CC diphosphate + L-methionyl-tRNA(Met).
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC MetG subfamily 1.
CC -1- SIMILARITY: Contains 1 tRNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE012861; AAM72204.1; -.
DR TIGR; CT0969; -.
DR HAMAP; MF_00098; fused; 1.
DR InterPro; IPR004495; MetG_Cterm.
DR InterPro; IPR008224; MetRS_dimerising.
DR InterPro; IPR008994; Nucleic acid OB.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002304; tRNA-synt_met.
DR InterPro; IPR002547; tRNA_bind.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR Pfam; PF01588; tRNA_bind; 1.
DR PIRSF; PIRSF001528; MeRS_dimerising; 1.
DR PRINTS; PR01041; TRNASYNTHMET.
DR TIGRFAMs; TIGR00398; metG; 1.
DR TIGRFAMs; TIGR00399; metG_Cterm; 1.
DR PROSITE; PS00178; AA_tRNA_LIGASE_1; 1.

```

```

DR PROSITE; PS00886; TRBD; 1.
KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KM RNA-binding; tRNA-binding; Metal-binding; Zinc; Complete proteome.
FT SITE 15 25 "HIGH" REGION.
FT SITE 345 349 "KMSKS" REGION.
FT DOMAIN 602 703 tRNA-BINDING.
FT METAL 147 147 ZINC (BY SIMILARITY).
FT METAL 150 150 ZINC (BY SIMILARITY).
FT METAL 160 160 ZINC (BY SIMILARITY).
FT METAL 163 163 ZINC (BY SIMILARITY).
FT BINDING 348 348 ATP (BY SIMILARITY).
SQ SEQUENCE 703 AA; 79553 MW; 9C4440919E7F52E6 CRC64;

Query Match 5.4%; Score 87.5; DB 1; Length 703;
Best Local Similarity 25.3%; Pred. No. 14;
Matches 40; Conservative 25; Mismatches 50; Indels 43; Gaps 9;

QY 125 DGDACRSRSPKQVELACGKSNRIAHVSEP-----STCVYALTFETPLVCHPH 172
DB 448 DGFHFRATQT-MEIA-RFANNFLTESEPMVKIKYDPEAGRTMAVSLN-----LCHTL 500
QY 173 ALVYPTLPALQRODQVE-QPLADELTTPQCHEKILTFEDAGYLTPEENPTQLE 231
DB 501 ALFWPIVPEPTANRIKMGFEGTIDLEVP-GNPVWRQAL-----EPLGKK 546
QY 232 G---GPDSTGFTELENCRKAHKEISKETRLKGLTQ 265
DB 547 GHKLSSSEILFSEKID-----KDIEPMKIKIALAE 579

```

Search completed: July 26, 2004, 11:09:46
 Job time : 7.08008 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:05:03 ; Search time 19.9295 Seconds

(without alignments)
4828.666 Million cell updates/sec

Title: US-10-657-280-3

Sequence: 1634
1 MAAGLARIILLILGLSAGSPA.....TPRAKSPQLRGDPGLRGSL 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL.25.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Print. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1634	100.0	305	4 Q9UJ39	Q9UJ39 homo sapien
2	1618.5	99.1	304	4 Q96L13	Q96L13 homo sapien
3	1601	98.0	331	4 Q96R22	Q96R22 homo sapien
4	1301.5	79.7	307	11 Q7TND0	Q7TND0 mus musculu
5	1288	78.8	314	11 Q8C5J3	Q8C5J3 mus musculu
6	185.5	11.4	507	5 Q23440	Q23440 caenorhabdi
7	172.5	10.6	529	13 Q80222	Q80222 brachydanio
8	171	10.5	647	10 Q9FM96	Q9FM96 arabidopsis
9	171	10.5	647	10 Q8LDD0	Q8LDD0 arabidopsis
10	155.5	9.5	612	4 Q95W99	Q95W99 homo sapien
11	155.5	9.5	615	4 Q81Z58	Q81Z58 homo sapien
12	148	9.1	548	5 Q9VUD1	Q9VUD1 drosophila
13	141	8.6	521	11 Q08795	Q08795 mus musculu
14	141	8.6	528	11 Q921X2	Q921X2 mus musculu
15	140.5	8.6	200	4 Q96BU9	Q96BU9 homo sapien
16	140.5	8.6	398	4 Q96D06	Q96D06 homo sapien

17	140.5	8.6	528	4 Q9P0W9	Q9P0W9 homo sapien
18	130	8.0	571	11 Q922S5	Q922S5 mus musculu
19	130	8.0	617	11 Q8K2C7	Q8K2C7 mus musculu
20	125	7.6	483	4 Q9NUY7	Q9NUY7 homo sapien
21	120.5	7.4	483	11 Q8VEH8	Q8VEH8 mus musculu
22	120.5	7.4	483	11 Q8BYN6	Q8BYN6 mus musculu
23	119	7.3	182	4 Q9UOL4	Q9UOL4 homo sapien
24	119	7.3	306	4 Q95901	Q95901 homo sapien
25	119	7.3	483	4 Q96D21	Q96D21 homo sapien
26	117	7.2	482	13 Q803G0	Q803G0 brachydanio
27	116.5	7.1	498	11 Q9CWA6	Q9CWA6 mus musculu
28	110.5	6.8	352	4 Q86WL7	Q86WL7 homo sapien
29	106.5	6.5	77	10 Q94AF1	Q94AF1 arabidopsis
30	106	6.5	150	10 Q65221	Q65221 arabidopsis
31	105.5	6.5	352	4 Q86VF1	Q86VF1 homo sapien
32	105.5	6.5	377	4 Q86U48	Q86U48 homo sapien
33	102.5	6.3	525	5 Q9VKC7	Q9VKC7 drosophila
34	100	6.1	534	5 Q9NFR2	Q9NFR2 caenorhabdi
35	98.5	6.0	216	12 Q91ZU7	Q91ZU7 chicken ane
36	96.5	5.9	216	12 Q91N82	Q91N82 chicken ane
37	96.5	5.8	216	12 Q9DXA8	Q9DXA8 chicken ane
38	95.5	5.8	216	12 Q8QY49	Q8QY49 chicken ane
39	95.5	5.8	216	12 Q95670	Q95670 chicken ane
40	95.5	5.8	216	12 Q91NA6	Q91NA6 chicken ane
41	95.5	5.8	282	10 Q8GWH3	Q8GWH3 arabidopsis
42	95.5	5.8	1839	2 Q30765	Q30765 streptomyce
43	95.5	5.8	4290	2 Q9WXC0	Q9WXC0 micromonos
44	95.5	5.8	4307	2 Q83WF0	Q83WF0 micromonos
45	95	5.8	407	16 Q66578	Q66578 aquifex aeo

ALIGNMENTS

RESULT 1

Q9UJ39 PRELIMINARY; PRT; 305 AA.

AC Q9UJ39;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE C316G12.3 (CAB56184).
GN C316G12.3 OR CAB56184.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Frankland J.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Fitzgerald P., Amarante-Vendes G.P., Li W., Green D.R.;
RT "cDNA from human fetal brain."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031709; CAB56184.1; -;
DR EMBL; AF02786; AAG27706.1; -;
DR PIR; T45062; T45062.
DR InterPro; IPR009011; Man_6_P_R_bind.
SQ SEQUENCE 305 AA; 33973 MW; 7774BBC091IDA1C2 CRC64;

Query Match 100.0%; Score 1634; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.2e-148;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARIILLILGLSAGSPAAGAAKCKVEEENAFGVNPFILPQASRLQAKRDPSPVG 60
DB 1 MAAGLARIILLILGLSAGSPAAGAAKCKVEEENAFGVNPFILPQASRLQAKRDPSPVG 60
QY 61 PVHLFRSLGKCELVSTYKFERCFPHNTQHEQTRRNVAISGILGIMHEWELNANTFFG 120

Db 61 PVLHRLSGKCSLVSTYKVFPCPHNVTOHEQTFRNNAVSGILGIWHEMIANPTFG 120
QY 121 MMARDGDACRSRSROSKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
Db 121 MMARDGDACRSRSROSKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALOROMQOVQODLADDELITPOGHEKILRTLPEDAGYLKTPPEENEPOLGSGDPSLGE 240
Db 181 PEALOROMQOVQODLADDELITPOGHEKILRTLPEDAGYLKTPPEENEPOLGSGDPSLGE 240
QY 241 TLENCKRAHKEISKEIKRLKGLITQHGIPYTRPETSNIHLGHETPRAKSPEDLRGDPG 300
Db 241 TLENCKRAHKEISKEIKRLKGLITQHGIPYTRPETSNIHLGHETPRAKSPEDLRGDPG 300
QY 301 LRGS 305
Db 301 LRGS 305
RESULT 2
Q96L13 PRELIMINARY; PRT; 304 AA.
ID Q96L13
AC Q96L13, 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC014592; AAK14592.1; -;
DR InterPro; IPR009011; Man_6_P_R_bind.
KW Hypothetical protein.
SQ SEQUENCE 304 AA; 33844 MW; FCIASA393DBB6C9 CRC64;

Query Match 99.1%; Score 1618.5; DB 4; Length 304;
Best Local Similarity 99.7%; Pred. No. 3.7e-147;
Matches 304; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MAGLARLLLLGLSAGPAPAGAAKMKVBEPPNAGVNNPFLPQASRLQAKRDPSPVSG 60
Db 1 MAGLARLLLLGLSAGPAPAGAAKMKVBEPPNAGVNNPFLPQASRLQAKRDPSPVSG 60
QY 61 PHLHRLSGKCSLVSTYKVFPCPHNVTOHEQTFRNNAVSGILGIWHEMIANPTFG 120
Db 61 PHLHRLSGKCSLVSTYKVFPCPHNVTOHEQTFRNNAVSGILGIWHEMIANPTFG 120
QY 121 MMARDGDACRSRSROSKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
Db 121 MMARDGDACRSRSROSKVELACGKSNRLAHVSEPTCYVALTFETPLVCHPHALLVPTL 180
QY 181 PEALOROMQOVQODLADDELITPOGHEKILRTLPEDAGYLKTPPEENEPOLGSGDPSLGE 240
Db 181 PEALOROMQOVQODLADDELITPOGHEKILRTLPEDAGYLKTPPEENEPOLGSGDPSLGE 240
QY 241 TLENCKRAHKEISKEIKRLKGLITQHGIPYTRPETSNIHLGHETPRAKSPEDLRGDPG 300
Db 241 TLENCKRAHKEISKEIKRLKGLITQHGIPYTRPETSNIHLGHETPRAKSPEDLRGDPG 300
QY 301 LRGS 305
Db 301 LRGS 305
RESULT 3
Q96R22 PRELIMINARY; PRT; 331 AA.
ID Q96R22

AC Q96R22;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to protein kinase C substrate.
GN RGD9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21096910; Pubmed=1157797;
RA Daniels R.J., Pedan J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tyfarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;
RT "Sequence, structure and pathology of the fully annotated terminal 2
KT kb of the short arm of human chromosome 16";
RL Hum. Mol. Genet. 10:339-352(2001).
DR EMBL; AE006467; AAK61277.1; -;
DR GO; GO:0016301; F.kinase activity; IEA.
DR InterPro; IPR009011; Man_6_P_R_bind.
KW Kinase.
SQ SEQUENCE 331 AA; 36799 MW; 0D01BBE475645F2F CRC64;

Query Match 98.0%; Score 1601; DB 4; Length 331;
Best Local Similarity 92.1%; Pred. No. 2e-145;
Matches 305; Conservative 0; Mismatches 0; Indels 26; Gaps 2;

QY 1 MAGLARLLLLGLSAGPAPAGAAKMKVBEPPNAGVNNPFLPQASRLQAKRDPSPVSG 59
Db 1 MAGLARLLLLGLSAGPAPAGAAKMKVBEPPNAGVNNPFLPQASRLQAKRDPSPVSG 60
QY 60 -----GPVHRLSGKCSLVSTYKVFPCPHNVTOHEQTFRNNAVSGI 104
Db 61 RFGAWTLGISVNSLPVHRLSGKCSLVSTYKVFPCPHNVTOHEQTFRNNAVSGI 120
QY 105 LGIWHMEIANTFTGMARDGDACRSRSROSKVELACGKSNRLAHVSEPTCYVALTPE 164
Db 121 LGIWHMEIANTFTGMARDGDACRSRSROSKVELACGKSNRLAHVSEPTCYVALTPE 180
QY 165 TPLVCHPHALLVPTLPALOROMQOVQODLADDELITPOGHEKILRTLPEDAGYLKTPPE 224
Db 181 TPLVCHPHALLVPTLPALOROMQOVQODLADDELITPOGHEKILRTLPEDAGYLKTPPE 240
QY 225 NEPTQEGDPSIGFETLNCRAHKEISKEIKRLKGLITQHGIPYTRP----- 274
Db 241 NEPTQEGDPSIGFETLNCRAHKEISKEIKRLKGLITQHGIPYTRPGEWMPGVLA 300
QY 275 ETSNLEHLGHETPRAKSPEDLRGDPG 305
Db 301 ETSNLEHLGHETPRAKSPEDLRGDPG 331

RESULT 4
Q7TNE0 PRELIMINARY; PRT; 307 AA.
ID Q7TNE0
AC Q7TNE0, 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=salivary gland;
RX MEDLINE=22388257; Pubmed=12477932;
RA Strausberg R.L., Reingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,

Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Ditchenko L., Marinova K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Tomshyuk S., Carrinci P., Prange C.,
 RA Raha S.S., Loguigliano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmitt J., Myers R.W., Buterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smaltus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Salivary gland;
 RA Strausberg R.;
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC055872; AAHS5872.1; --
 DR Hypothetical protein.
 SQ SEQUENCE 307 AA; 34109 MW; 8712A9E4288BD56 CRC64;

Query Match 79.7%; Score 1301.5; DB 11; Length 307;
 Best Local Similarity 78.8%; Pred. No. 1.2e-116;
 Matches 241; Conservative 25; Mismatches 39; Indels 1; Gaps 1;

QY 1 MAAGIARLLILIGLSAGGPAPAGAAKMYEEPNAGVNNPFLPQASRLQKRPSPVSG 60
 DB 1 MAGRLAGFLMLGLASQGPAPACAGKMYEEPNFTGLNPPFLPQASRLQKRPSPVSG 60
 QY 61 PHLFRLSGKCFSLVSTYKFCPEPHNTQHEQTFRNAYSGIIGIWHMEIANNTFTG 120
 DB 61 PHLFRLSGKCFSLVSTYKFCPEPHNTQHEQTFRNAYSGIIGIWHMEIANNTFTG 120
 QY 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPHALVPTL 180
 DB 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPHALVPTL 180
 QY 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPHALVPTL 180
 DB 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPHALVPTL 180
 QY 181 PEALGRQNDQVQDADLDELITPGHEKILRTLFEDAGYLKTPBESEPTQLGSGDSGFE 240
 DB 181 PEALGRQNDQVQDADLDELITPGHEKILRTLFEDAGYLKTPBESEPTQLGSGDSGFE 240
 QY 241 TLNCRKAHKEISKEIKLKGILLTQHGIPYTPETETSNLEHGHETP-RAKSPQLRGDP 299
 DB 241 TLNCRKAHKEISKEIKLKGILLTQHGIPYTPETETSNLEHGHETP-RAKSPQLRGDP 299
 QY 241 TLNCRKAHKEISKEIKLKGILLTQHGIPYTPETETSNLEHGHETP-RAKSPQLRGDP 299
 DB 241 TLNCRKAHKEISKEIKLKGILLTQHGIPYTPETETSNLEHGHETP-RAKSPQLRGDP 299
 QY 300 GLRGSL 305
 DB 301 GLRGSL 306

RESULT 5
 Q8C5J3 PRELIMINARY; PRT; 314 AA.
 ID Q8C5J3
 AC Q8C5J3;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
 DE Hypothetical extracytoplasmic domain of cation-dependent mannose
 DE 6-phosphate receptor structure containing protein.
 GN 130004G08RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=22354683; PubMed=1246851;

RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL: AK078230; BAC37183.1; --
 DR MGD: 1921404; 130004G08RIK.
 DR InterPro: IPR009011; Man_6_P_R_bind.
 KW Hypothetical protein.
 SQ SEQUENCE 314 AA; 34891 MW; 7C3424A15649D058 CRC64;

Query Match 78.8%; Score 1288; DB 11; Length 314;
 Best Local Similarity 77.0%; Pred. No. 2.4e-115;
 Matches 241; Conservative 25; Mismatches 39; Indels 8; Gaps 2;

QY 1 MAAGIARLLILIGLSAGGPAPAGAAKMYEEPNAGVNNPFLPQASRLQKRPSPVSG 60
 DB 1 MAGRLAGFLMLGLASQGPAPACAGKMYEEPNFTGLNPPFLPQASRLQKRPSPVSG 60
 QY 61 PHLFRLSGKCFSLVSTYKFCPEPHNTQHEQTFRNAYSGIIGIWHMEIANNTFTG 120
 DB 61 PHLFRLSGKCFSLVSTYKFCPEPHNTQHEQTFRNAYSGIIGIWHMEIANNTFTG 120
 QY 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPHALVPTL 173
 DB 121 MMNRDGDACRSRQSKYELACGKSNRLAHVSEPTCYALTFTPLVCHPHALVPTL 180
 QY 174 LLYVPTLPEALGRQNDQVQDADLDELITPGHEKILRTLFEDAGYLKTPBESEPTQLGSG 233
 DB 181 LLYVPTLPEALGRQNDQVQDADLDELITPGHEKILRTLFEDAGYLKTPBESEPTQLGSG 240
 QY 234 PDSLGFETLENCRKAHKEISKEIKLKGILLTQHGIPYTPETETSNLEHGHETP-RAKSP 292
 DB 241 SKELGELTLNCRKAHKEISKEIKLKGILLTQHGIPYTPETETSNLEHGHETP-RAKSP 300
 QY 293 EQLRGDPGLRGSL 305
 DB 301 EHLRSDPGLRGSL 313

RESULT 6
 Q23440 PRELIMINARY; PRT; 507 AA.
 ID Q23440
 AC Q23440; Q22519;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
 DE ZK1307.8 protein.
 GN ZK1307.8
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peleiderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Berks M.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z47356; CAAB7420.1; --
 DR EMBL: Z47358; CAAB7420.1; JOINED.
 DR EMBL: Z47358; CAAB7438.1; --
 DR EMBL: Z47356; CAAB7438.1; JOINED.
 DR PIR: T24944; T24944.
 DR WormPep: ZK1307.8; CB15547.
 DR GO: 0005509; F:calcium ion binding; IEA.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR00086; EF-hand.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR009011; Man_6_P_R_bind.
 DR Pfam: PF00036; ehand; 1.
 DR SMART: SM00192; LDpa; 1.
 DR PROSITE: PS00018; EF_HAND; 2.
 DR PROSITE: PS00014; ER_TARGET; 1.
 SQ SEQUENCE 507 AA; 58062 MW; 2764EC56741227DF CRC64;

Db 536 KTYVVCAYKATQBE-----GYSKTRLGEMDKFE--NSQWSTYNKCKMNGPDRS 586
QY 136 SKVELACCKSNRLAHVSEPTCYVALTFETPLVCHPHALLVYPTLPALQROMDQVEODL 195
Db 587 LKVKLRGCKNKLMDVDDEPSRCEVAAALISTPARC-----LEDXKLKEIQLEKLMNQD- 639
QY 196 ADELITPQCHEKL 208
Db 640 -----KQNMDEL 647

RESULT 10
Q9BM99
ID Q9BM99, PRELIMINARY; PRT; 612 AA.
AC Q9BM99, 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to amplified in osteosarcoma.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC000532; AAH00532.1; -
DR InterPro; IPR009011; Man_6_P_R_bind.
SQ SEQUENCE 612 AA; 69246 MW; 1FD06F7B23101015 CRC64;

Query Match 9.5%; Score 155.5; DB 4; Length 612;
Best Local Similarity 21.7%; Pred. No. 5.5e-06;
Matches 79; Conservative 47; Mismatches 121; Indels 117; Gaps 18;

QY 21 PAGAAMKVVVEEPNAGVNNPFLPQASRLQAKRDPSPVSGVHLFRLSGKCFSLVESTYK 80
Db 76 PAGAHHFQREBERETPAVQGPGLPDL--LSPMRD-----APCLTKTKDMWT 119
QY 81 YEFCEPHNVYQ-----HEQTFRNNAVSGILGIWHEWEIANNFTGMM 123
Db 120 YEFCEGRHQQYHMEDESEIKGEVLYGYQSAFMDDEDTAKASKQRLKRYHSQTYG--- 176
QY 124 RQGDAC--RSRSQSVLELACGK-----SNRLAHVSEPTCYVALTFETPLVC----- 169
Db 177 -NGSKCDLNGRPREAEVRFLODEGAGISGDYIDRVDEPLSCSYVLTITPRLCPHPLRP 235
QY 170 ----HPHALVYPTL-PEA---LQROMDQVE-----ODLADL-----ITPQ- 203
Db 236 PPSAAPQALICHPSLOPEEYMAVYQROADSKQYGDKIIEELQDLGQVWSETKSGVAPQK 295
QY 204 -----GHEKILRLTFEDA-GYLKTP-EENEPT-----QLEGGPDLSGFETL 242
Db 296 MAGASPTKDDSKDQSPWKLNEPEEQAPGGEVPAEEDQPSPAADSAGAPYDFQNNVQ 355
QY 243 ENCRKAHKEISKELKGLLTQHGIPYTRPETSLSLHLSHETP-----RAKSPEDL 295
Db 356 VAVIISPADLIRFIEELKG-GTKKGP-----NIGQEQVDDAAEVAPQREPEKE 403
QY 296 RQDP 299
Db 404 RQDP 407

RESULT 11
Q81Z58
ID Q81Z58, PRELIMINARY; PRT; 615 AA.
AC Q81Z58, 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (Feb-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC023513; AAH23513.1; -
DR InterPro; IPR009011; Man_6_P_R_bind.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 615 AA; 69641 MW; 757D625B6FDA7097 CRC64;

Query Match 9.5%; Score 155.5; DB 4; Length 615;
Best Local Similarity 21.7%; Pred. No. 5.5e-06;
Matches 79; Conservative 47; Mismatches 121; Indels 117; Gaps 18;

QY 21 PAGAAMKVVVEEPNAGVNNPFLPQASRLQAKRDPSPVSGVHLFRLSGKCFSLVESTYK 80
Db 79 PAGAHHFQREBERETPAVQGPGLPDL--LSPMRD-----APCLTKTKDMWT 122
QY 81 YEFCEPHNVYQ-----HEQTFRNNAVSGILGIWHEWEIANNFTGMM 123
Db 123 YEFCEGRHQQYHMEDESEIKGEVLYGYQSAFMDDEDTAKASKQRLKRYHSQTYG--- 179
QY 124 RQGDAC--RSRSQSVLELACGK-----SNRLAHVSEPTCYVALTFETPLVC----- 169
Db 180 -NGSKCDLNGRPREAEVRFLODEGAGISGDYIDRVDEPLSCSYVLTITPRLCPHPLRP 238
QY 170 ----HPHALVYPTL-PEA---LQROMDQVE-----ODLADL-----ITPQ- 203
Db 239 PPSAAPQALICHPSLOPEEYMAVYQROADSKQYGDKIIEELQDLGQVWSETKSGVAPQK 298
QY 204 -----GHEKILRLTFEDA-GYLKTP-EENEPT-----QLEGGPDLSGFETL 242
Db 299 MAGASPTKDDSKDQSPWKLNEPEEQAPGGEVPAEEDQPSPAADSAGAPYDFQNNVQ 358
QY 243 ENCRKAHKEISKELKGLLTQHGIPYTRPETSLSLHLSHETP-----RAKSPEDL 295
Db 359 VAVIISPADLIRFIEELKG-GTKKGP-----NIGQEQVDDAAEVAPQREPEKE 406
QY 296 RQDP 299
Db 407 RQDP 410

RESULT 12
Q9VUD1
ID Q9VUD1, PRELIMINARY; PRT; 548 AA.
AC Q9VUD1, 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE C66453 protein (LD46533P).
GN C66453.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Cainlaker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell W.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

Query	DB	Score	Length	Matches	Indels	Gaps
60	GPVH.FR-1JSGCFLVSTYKYEFCFPHNTQHOEOTFRNNAYSG--ILGIHWE-1A	114	521	36	15	5
396	GGSGEPAYIXSOCLITLNEVYALCPFKLYSQPK-----HGSSPTSLGTGWSAGD	449	505	36	15	5
115	NNFTFGMMNRDDAC-RSRSRQSGKVELACGSKNLAHVSBPTCYVALTFEPPLVC	169	505	36	15	5
450	HDKFSAMKYEQGTGCGQPNRSTTVRLCGXETVVTSTPSPRCEYIMELMTPAAC	505	505	36	15	5

DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR009011; Man_6_P_R_bind.
 DR Pfam; PF00036; ehand; 2.
 DR SMART; SM00192; LDla; 1.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 SQ SEQUENCE 528 AA; 59559 MW; 253FDIA76A641713 CRC64;

Query Match 8.6%; Score 141; DB 11; Length 528;
 Best Local Similarity 31.0%; Pred. No. 0.00011;
 Matches 36; Conservative 15; Mismatches 53; Indels 12; Gaps 5;

QY 60 GPVHLFR-LSGKCFSLVESTYKYEFCFPHNVTQHEQTFRMNAYSG--ILGIHWE-IA 114
 DB 403 GPSGFEALYSGCYELTNEVYRLCPFKLYSQKPK-----HGSPISLGTGSMAGPD 456
 QY 115 NNTFTGMMWRDGDAC-RSRSRQSKVELACGKSNRLAHVSEPSCTCYALTFFETPLVC 169
 DB 457 HDKFSAMKYEQGTGCGPNRSTVRLCGKETVVTSTTEPSRCEYLMELMTPAAC 512

RESULT 15

ID Q96BU9 PRELIMINARY; PRT; 200 AA.
 AC Q96BU9;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC015154; AAH15154.1; -.
 DR InterPro; IPR000886; ER_target_S.
 DR InterPro; IPR009011; Man_6_P_R_bind.
 DR PROSITE; PS00014; ER_TARGET; 1.
 KM Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 200 AA; 22628 MW; B9PA34051267DE9E CRC64;

Query Match 8.6%; Score 140.5; DB 4; Length 200;
 Best Local Similarity 30.5%; Pred. No. 3.3e-05;
 Matches 32; Conservative 14; Mismatches 54; Indels 5; Gaps 3;

QY 67 LSGKCFSLVESTYKYEFCFPHNVTQHEQTFRMNAYSGILGIHWE-ETANTFTGMMWRD 125
 DB 83 LYSQCYELTNEVYRLCPFKLYSQKPK---KLGSPTSLGTGSMWIGPDHDKFSAMKYEQ 139
 QY 126 GDAC-RSRSRQSKVELACGKSNRLAHVSEPSCTCYALTFFETPLVC 169
 DB 140 GTGCMQGPNRSTVRLCGKETVVTSTTEPSRCEYLMELMTPAAC 184

Search completed: July 26, 2004, 11:11:40
 Job time : 22.925 secs

This Page Blank (uspto)